Parametric first-order Edgeworth expansion for Markov additive functionals. Application to M-estimations.

Déborah Ferré

Parametric first-order Edgeworth expansion for Markov additive functionals. Application to M-estimations.

D.FERRE *

février 2012

Abstract

We give a spectral approach to prove a parametric first-order Edgeworth expansion for bivariate additive functionals of strongly ergodic Markov chains. In particular, given any V-geometrically ergodic Markov chain $(X_n)_{n\in\mathbb{N}}$ whose distribution depends on a parameter θ, we prove that $\{\xi_p(X_{n-1}, X_n);\ p \in \mathcal{P},\ n \geq 1\}$ satisfies a uniform (in (θ, p)) first-order Edgeworth expansion provided that $\{\xi_p(\cdot,\cdot);\ p \in \mathcal{P}\}$ satisfies some non-lattice condition and an almost optimal moment domination condition. Furthermore, the sequence $(X_n)_{n\in\mathbb{N}}$ need not be stationary. This result is applied to M-estimations.

1 Introduction

Let $(E, \mathcal{E})$ be any measurable space, and let $(X_n)_{n\geq 0}$ be a Markov chain on a general state space E with transition kernel $(Q_\theta(x,\cdot);\ x \in E)$ where θ is a parameter in some set Θ. The initial distribution of the chain is denoted by μ_θ. The underlying probability measure is denoted by $\mathbb{P}_{\theta,\mu_\theta}$.

Let $\{\xi_p(\cdot,\cdot);\ p \in \mathcal{P}\}$ be a family of measurable functions from E^2 into $\mathbb{R}$, where $\mathcal{P}$ is any set. Let us define the following bivariate additive functionals

$$\forall n \geq 1,\ \forall p \in \mathcal{P}, \qquad S_n(p) := \sum_{k=1}^{n} \xi_p(X_{k-1}, X_k). \tag{1}$$

We are interested in appropriate conditions on the model, on the family $\{\xi_p(\cdot,\cdot);\ p \in \mathcal{P}\}$ and on the initial probability measure μ_θ, under which a first-order Edgeworth expansion exists (also called Esseen theorem), namely there exists a polynomial function $A_{\theta,p}(\cdot)$ such that

$$\sup_{(\theta,p)\in\Theta\times\mathcal{P}} \sup_{u\in\mathbb{R}} \left| \mathbb{P}_{\theta,\mu_\theta}\left\{ \frac{S_n(p)}{\sigma_{\theta,p}\sqrt{n}} \leq u \right\} - \mathcal{N}(u) - \eta(u)\ n^{-\frac{1}{2}}\ A_{\theta,p}(u) \right| = o(n^{-\frac{1}{2}}), \tag{2}$$

*Université Européenne de Bretagne, I.R.M.A.R. (UMR-CNRS 6625), Institut National des Sciences Appliquées de Rennes. Deborah.Ferre@insa-rennes.fr

where $\mathcal{N}$ is the standard normal distribution function and η is its density. Note that Expansion (2) holds uniformly in $(\theta, p) \in \Theta \times \mathcal{P}$. As illustrated later in M-estimation, the bivariate and parametric form of (1), as well as the previous uniform control and the possible non-stationarity of μ_θ, are required for statistical applications.

Edgeworth expansions in the Markov setting can be established by the two following methods:

1. *The regeneration method.* This standard method, introduced by [Smi55], was used by Bolthausen [Bol82] to establish the Berry-Esseen theorem for univariate additive functionals of the form $S_n = \sum_{k=1}^n \xi(X_k)$, by splitting S_n into a sum of independent blocks. This method can be applied to the general class of Harris-recurrent chains $(X_n)_{n\geq 0}$ which either possess an accessible atom or satisfy some minorization condition. Bolthausen work was extended to Edgeworth expansions by Malinovskii [Mal87] and next generalized to bivariate additive functionals $S_n = \sum_{k=1}^n \xi(X_{k-1}, X_k)$ by Jensen [Jen89].
   Note that in [Bol82, Mal87, Jen89], neither the distribution of $(X_n)_{n\geq 0}$ nor the function ξ depends on parameters. However a recent work due to Bertail-Clémençon [BC11] provides a Berry-Esseen theorem adapted to the above mentioned parametric setting, but the extension to Edgeworth expansions would generate even more difficulties. Furthermore this statement only concerns univariate additive functionals and the extension of their proof to the bivariate case (1) induces dependence between the regeneration blocks and hence provides at least one more difficulty to handle with. For further explanations concerning extension of Bertail-Clémençon work, see Appendix B.
   Moreover all these works are valid under some complex block-moment conditions. If the strong mixing coefficient of $(X_n)_{n\geq 0}$ decreases at a fast enough rate, then these block-moment conditions are entailed by some explicit moment condition provided that the initial probability is dominated by some multiple of the stationary distribution π. When considering the particular case where the chain possesses an atom A, this simplification also holds true whenever the initial probability μ is the Dirac distribution δ_x at some $x \in A$. Let us note that the resulting moment condition is then almost optimal (with respect to the independent case). However, to the best of our knowledge, this simplification cannot be extended to the case where the initial probability μ is any probability distribution, in particular where μ is the Dirac distribution δ_x at some $x \in E$ which does not necessarily belong to an atom. For further explanations concerning conditions on μ, see Appendix B.

2. *The weak Nagaev-Guivarc'h spectral method.* This method, based on the Keller-Liverani perturbation theorem [KL99], enables the statement of limit theorems for additive functionals associated to strongly ergodic Markov chains (Harris recurrence is no more required). This method has been fully described in [HP10] in the case of univariate additive functionals. It is specially efficient for ρ-mixing and V-geometrically Markov chains, as well as for iterated function systems. In those models, the extension of Berry-Esseen type results of [HP10] to the case of bivariate additive functionals of the type (1) has already been obtained in [FHL, HLP][1] with in addition the desired control on the parameters

[1] and in the paper intitled "Regularity of the characteristic function of additive functionals for iterated function systems. Statistical applications"', which is to be submitted very soon (authors: D.Guibourg and D.Ferré).

(θ, p). The resulting moment conditions on $\{\xi_p;\ p \in \mathcal{P}\}$ are (almost) optimal with respect to the independent case. Let us note that they are explicit for these three models and do not depend on the initial probability (unlike the ones given by the regeneration method).

In this paper, we will state Expansion (2) when concerning with the class of strongly ergodic Markov chains, and apply Fourier techniques via the perturbation operator theory of Nagaev-Guivarc'h.

Our work extends the Berry-Esseen type results of [HLP] to the first-order Edgeworth expansion. As in the independent case, the gap from Berry-Esseen to Edgeworth type results induces at least a new difficulty: the requirement of the non-arithmeticity hypothesis.

In Subsection 2.1, we consider a family of random variables (r.v.) $S_n(p)$ (not necessarily derived from Markovian models) defined on a general parametric probabilistic space $(\Omega, \mathcal{F}, \{\mathbb{P}_\theta; \theta \in \Theta\})$, and we state hypotheses called $\mathcal{R}(m)$ and (N-A) under which Expansion (2) holds true. These hypotheses concern the behavior of the characteristic function $t \mapsto \phi_{n,p}(t)$ of $S_n(p)$: Hypothesis $\mathcal{R}(m)$ focuses on the form and the regularity of $\phi_{n,p}$ near $t = 0$; whereas Hypothesis (N-A), related to the non-arithmeticity assumption, focuses on the behavior of $\phi_{n,p}$ outside $t = 0$.

In Subsection 2.2, we specify the form of $S_n(p)$: from now on, $S_n(p)$ is defined by (1) where $(X_n)_{n\geq 0}$ is assumed to be a strongly ergodic Markov chain, and we give a brief review of the weak Nagaev-Guivarc'h spectral method to check Hypothesis $\mathcal{R}(m)$ and (N-A) in this Markov context. In fact, as already done in [FHL, HLP], Hypothesis $\mathcal{R}(m)$ can be investigated thanks to an easy extension of the results of [HP10].

By contrast, the method developed in [HP10] is not sufficient to study Hypothesis (N-A). Indeed, the non-arithmeticity condition has to be checked uniformly in both the parameter θ of the Markovian model and the parameter p of the family $\{\xi_p;\ p \in \mathcal{P}\}$ involved in (1). The study of (N-A) in this context is original and constitutes one of the main contributions of this paper (actually, even in the independent case, this question is far from being obvious). In our Markov setting, this study is based on the operator perturbation theory, quasi-compactness arguments and Ascoli theorem. Specifically, in Section 3, we give three approaches to reduce Hypothesis (N-A) to some simple non-lattice conditions in the case of general strongly ergodic Markov chains.

Section 4 is devoted to V-geometrically ergodic Markov chains. For this instance and more specifically for dominated models, we reduce (N-A) using one of the three approaches presented in Subsection 3.3. Combining this result together with the sufficient conditions of [HLP] to check Hypothesis $\mathcal{R}(m)$ and the general Edgeworth type statement of Subsection 2.1, provides Expansion (2) under assumptions close to the ones of the independent case.

Statistical applications are studied in Section 5: a first-order Edgeworth expansion for M-estimators of V-geometrically ergodic Markov chains is derived (Theorem 2) from the results of Section 4. Theorem 2, which extends Pfanzagl theorem [Pfa73] obtained for independent and identically distributed (i.i.d.) data under some moment conditions of order 3, is valid under a natural adaptation of the statistical regularity conditions of [Pfa73], moment domination conditions of order $3 + \varepsilon$, and some simple non-lattice condition as well. To the best of our knowledge, this result is new. Notice that our moment domination conditions are not only almost optimal, but also take the same form as the ones used in [DY07] to prove the asymptotic normality of M-estimators under V-geometrically ergodicity. This result is illustrated with AR(1) processes in Subsection 5.2.

The adaptation of Pfanzagl proof is developed in Section 6 for general statistical models under Hypotheses $\mathcal{R}(3)$ and (N-A). Note that this adaptation is not straightforward. Finally, the intermediate results of this section are applied in Subsection 6.4 to M-estimators of AR(d) processes.

2 Fourier techniques and first-order Edgeworth expansion

In this section, we present some results based on Fourier techniques. These results appeal to the next Hypotheses $\mathcal{R}(m)$ and (N-A) that are well-suited for the markovian case as explained in Subsection 2.2.

2.1 Hypotheses $\mathcal{R}(m)$ and (N-A) and first-order Edgeworth expansion

Let $(\Omega, \mathcal{F}, \{\mathbb{P}_\theta; \theta \in \Theta\})$ be any statistical model, where Θ is some parameter space. The underlying expectation is denoted by $\mathbb{E}_\theta$. Consider a family $\{S_n(p);\ n \in \mathbb{N}^*,\ p \in \mathcal{P}\}$ of real r.v. defined on $(\Omega, \mathcal{F}, \{\mathbb{P}_\theta; \theta \in \Theta\})$, where $\mathcal{P}$ is any set. Note that the parameter p may depend on θ.

Hypothesis $\mathcal{R}(m)$, $m \in \mathbb{N}^*$. *There exists a bounded open interval $I_0 \subset \mathbb{R}$ of $t = 0$ such that one has for all $(\theta, p) \in \Theta \times \mathcal{P}$, $n \geq 1$, $t \in I_0$*

$$\mathbb{E}_\theta[e^{itS_n(p)}] = \lambda_{\theta,p}(t)^n \left(1 + l_{\theta,p}(t)\right) + r_{\theta,p,n}(t), \tag{3}$$

where $\lambda_{\theta,p}(\cdot)$, $l_{\theta,p}(\cdot)$ and $r_{\theta,p,n}(\cdot)$ are $\mathbb{C}$-valued functions of class $\mathcal{C}^m$ on I_0 satisfying the following properties:

$$\lambda_{\theta,p}(0) = 1, \quad \lambda^{(1)}_{\theta,p}(0) = 0, \quad l_{\theta,p}(0) = 0, \quad r_{\theta,p,n}(0) = 0,$$

and for $\ell = 0, \ldots, m$

$$\sup\left\{|\lambda^{(\ell)}_{\theta,p}(t)|;\ t \in I_0,\ (\theta,p) \in \Theta \times \mathcal{P}\right\} < +\infty,$$

$$\sup\left\{|l^{(\ell)}_{\theta,p}(t)|;\ t \in I_0,\ (\theta,p) \in \Theta \times \mathcal{P}\right\} < +\infty,$$

$$\exists \kappa \in [0,1),\ \exists G_\ell > 0,\ \forall n \geq 1, \quad \sup\left\{|r^{(\ell)}_{\theta,p,n}(t)|; t \in I_0,\ (\theta,p) \in \Theta \times \mathcal{P}\right\} \leq G_\ell\, \kappa^n.$$

Furthermore, the functions $\lambda^{(m)}_{\theta,p}(\cdot)$, $l^{(m)}_{\theta,p}(\cdot)$ and $r^{(m)}_{\theta,p,n}(\cdot)$ are continuous on I_{00} uniformly in $(\theta, p) \in \Theta \times \mathcal{P}$.

Hypothesis (N-A) (Non-arithmeticity). *For any compact subset K_0 of $I\!R^*$, there exists $\rho \in [0,1)$ such that*

$$\forall n \geq 1, \quad \sup \left\{ \left| \mathbb{E}_\theta[e^{itS_n(p)}] \right|;\ t \in K_0,\ (\theta,p) \in \Theta \times \mathcal{P} \right\} = O(\rho^n).$$

Note that under Hypothesis $\mathcal{R}(2)$, the function $t \mapsto \mathbb{E}_\theta[e^{itS_n(p)}]$ is of class $\mathcal{C}^2$ on I_0 for all $(\theta,p) \in \Theta \times \mathcal{P}$. Then by Fatou lemma, for all $(\theta,p) \in \Theta \times \mathcal{P}$, one has $\mathbb{E}_\theta[S_n(p)^2] < +\infty$. Therefore, when considering the derivative of Equality (3), one easily obtains that for all $(\theta,p) \in \Theta \times \mathcal{P}$, $\lim \mathbb{E}_\theta[S_n(p)]/n = 0$ when $n \to +\infty$. Note that under Hypothesis $\mathcal{R}(2)$, when considering the derivative of Equality (3), one easily obtains as well

$$\forall n \geq 1, \quad \lim_{n \to +\infty} \sup_{(\theta,p) \in \Theta \times \mathcal{P}} \left| \frac{\mathbb{E}_\theta[S_n(p)^2]}{n} \right| < +\infty, \tag{4}$$

and in a similar way, under Hypothesis $\mathcal{R}(4)$,

$$\forall n \geq 1, \quad \lim_{n \to +\infty} \sup_{(\theta,p) \in \Theta \times \mathcal{P}} \left| \frac{\mathbb{E}_\theta[S_n(p)^4]}{n^2} \right| < +\infty. \tag{5}$$

Finally, under Hypothesis $\mathcal{R}(3)$, we obtain some of the assertions of Proposition 1 below. The other ones can be proved by borrowing the proof of [Fel71].

Proposition 1 (first-order Edgeworth expansion).
If $\{S_n(p);\ n \in I\!N^*,\ p \in \mathcal{P}\}$ satisfies Hypothesis $\mathcal{R}(3)$, then for all $(\theta,p) \in \Theta \times \mathcal{P}$, the following limits

$$b_{\theta,p} := \lim_{n \to +\infty} \mathbb{E}_\theta[S_n(p)], \qquad \sigma^2_{\theta,p} := \lim_{n \to +\infty} \frac{\mathbb{E}_\theta[S_n(p)^2]}{n},$$

are well-defined and bounded in $\theta \in \Theta$. Furthermore if $\inf_{(\theta,p) \in \Theta \times \mathcal{P}} \sigma_{\theta,p} > 0$ and if the family $\{S_n(p);\ n \in I\!N^*,\ p \in \mathcal{P}\}$ satisfies Hypothesis (N-A) as well, then there exists a polynomial function $G_{n,\theta,p}$ such that

$$\sup_{(\theta,p) \in \Theta \times \mathcal{P}} \sup_{u \in I\!R} \left| \mathbb{P}_\theta \left\{ \frac{S_n(p)}{\sigma_{\theta,p}\sqrt{n}} \leq u \right\} - G_{n,\theta,p}(u) \right| = o(n^{-\frac{1}{2}}).$$

The polynomial function $G_{n,\theta,p}$ is of the type $G_{n,\theta,p}(u) = \mathcal{N}(u) + \left[a_1(\theta,p) + a_2(\theta,p)\, u^2\right]\, \eta(u)/\sqrt{n}$ where the coefficients satisfy for $i = 1,2$, $\sup_{(\theta,p) \in \Theta \times \mathcal{P}} |a_i(\theta,p)| < +\infty$. Furthermore, if $\mathbb{E}_\theta[|S_n(p)|^3] < +\infty$ for all $n \geq 1$ and $(\theta,p) \in \Theta \times \mathcal{P}$, then the limit

$$m^3_{\theta,p,3} := \lim_{n \to +\infty} \frac{\mathbb{E}_\theta[S_n(p)^3]}{n} - 3\sigma^2_{\theta,p} b_{\theta,p}$$

is well-defined and bounded in $\theta \in \Theta$, and moreover the polynomial function $G_{n,\theta,p}$ can be explicitly expressed by

$$G_{n,\theta,p}(u) := \mathcal{N}(u) + \frac{m^3_{\theta,p,3}}{6\sigma^3_{\theta,p}\sqrt{n}}(1-u^2)\,\eta(u) - \frac{b_{\theta,p}}{\sigma_{\theta,p}\sqrt{n}}\,\eta(u).$$

Remark 1. *In the i.i.d. case, Hypotheses $\mathcal{R}(3)$ and (N-A) are easily checked. Indeed consider $(X_n)_{n\in\mathbb{N}^*}$ a sequence of i.i.d. E-valued r.v. whose common distribution depends on $\theta \in \Theta$, and $\{\xi_p(\cdot);\, p \in \mathcal{P}\}$ a family of measurable functions from E into $\mathbb{R}$. The following assertions are obviously equivalent:*

(a) The family $\{\sum_{k=1}^n \xi_p(X_k);\ n \in \mathbb{N}^*,\ p \in \mathcal{P}\}$ fulfills Hypothesis $\mathcal{R}(m)$ if and only if for all $(\theta, p) \in \Theta \times \mathcal{P}$, $\mathbb{E}_\theta[\xi_p(X_1)] = 0$, and $\sup_{(\theta,p)\in\Theta\times\mathcal{P}} \mathbb{E}_\theta[|\xi_p(X_1)|^m] < +\infty$.

(b) The family $\{\sum_{k=1}^n \xi_p(X_k);\ n \in \mathbb{N}^*,\ p \in \mathcal{P}\}$ fulfills Hypothesis (N-A) if and only if, for any compact subset K_0 of $\mathbb{R}^*$, one has

$$\sup_{t\in K_0} \sup_{(\theta,p)\in\Theta\times\mathcal{P}} \left|\mathbb{E}_\theta[e^{it\xi_p(X_1)}]\right| < 1. \tag{6}$$

When (6) is considered at (θ, p) fixed, it can be easily relaxed to the usual condition: $\xi_p(X_1)$ is non-lattice. By contrast, it is not easy to relax the uniform condition (6). Note that this condition is only discussed in [Pfa73] under the stronger Cramér condition:

$$\limsup_{t\to+\infty} \sup_{(\theta,p)\in\Theta\times\mathcal{P}} \left|\mathbb{E}_\theta[e^{it\xi_p(X_1)}]\right| < 1.$$

Hypotheses $\mathcal{R}(m)$ and (N-A) are the tailor-made assumptions to borrow the proof of the first-order Edgeworth expansion in the i.i.d. case [2], and consequently to expand $\mathbb{P}_\theta\{S_n(p)/(\sigma_{\theta,p}\sqrt{n}) \le u\}$ with a polynomial function independent on n. Notice that, under less restrictive conditions, the results of [Dur80] provide a first-order Edgeworth-type expansion but with a polynomial function depending on n.

2.2 The main lines of the weak spectral method for Markovian models

Consider from now on the following general Markovian setting. Let $(E, \mathcal{E})$ be any measurable space, and let $(X_n)_{n\ge0}$ be a Markov chain with state space E and transition kernel $(Q_\theta(x,\cdot);\ x \in E)$ where θ is a parameter in some set Θ. The initial distribution of the chain is denoted by μ_θ (i.e. $X_0 \sim \mu_\theta$). The underlying probability measure and the associated expectation are denoted by $\mathbb{P}_{\theta,\mu_\theta}$ and $\mathbb{E}_{\theta,\mu_\theta}$. We assume that $(X_n)_{n\in\mathbb{N}}$ admits an invariant probability measure denoted by π_θ (i.e. $\forall\theta \in \Theta$, $\pi_\theta \circ Q_\theta = \pi_\theta$). Notice that we do not require stationarity for $(X_n)_{n\in\mathbb{N}}$.

Let $\{\xi_p(\cdot,\cdot);\, p \in \mathcal{P}\}$ be a family of measurable functions from E^2 into $\mathbb{R}$, where $\mathcal{P}$ is any set. Let us define the following r.v.

$$\forall n \ge 1,\ \forall p \in \mathcal{P}, \qquad S_n(p) := \sum_{k=1}^n \xi_p(X_{k-1}, X_k). \tag{7}$$

This kind of (parametric and bivariate) functionals is required when concerning with Markovian M-estimators, as detailed in Section 5.

[2] One difference is that the asymptotic bias $b_{\theta,p} = 0$ in the i.i.d. case.

Now we are going to study Hypotheses $\mathcal{R}(m)$ and (N-A) using the Nagaev-Guivarc'h spectral method. For all $t \in I\!R$, $(\theta, p) \in \Theta \times \mathcal{P}$ and $x \in E$, let us define the Fourier kernel of (Q_θ, ξ_p) by

$$Q_{\theta,p}(t)(x, dy) := e^{it\xi_p(x,y)} Q_\theta(x, dy). \tag{8}$$

As usual, for all bounded measurable $\mathbb{C}$-valued function f on E, we set

$$Q_{\theta,p}(t)f := \int_E f(y) e^{it\xi_p(\cdot,y)} Q_\theta(\cdot, dy).$$

It is easy to see that we have from Markov property

$$\forall t \in I\!R,\ \forall (\theta, p) \in \Theta \times \mathcal{P},\ \forall n \geq 1, \qquad \mathbb{E}_{\theta,\mu_\theta}[e^{itS_n(p)} f(X_n)] = \mu_\theta[Q_{\theta,p}(t)^n f].$$

In particular, we obtain

$$\forall t \in I\!R,\ \forall (\theta, p) \in \Theta \times \mathcal{P},\ \forall n \geq 1, \qquad \mathbb{E}_{\theta,\mu_\theta}[e^{itS_n(p)}] = \mu_\theta[Q_{\theta,p}(t)^n \mathbf{1}_E], \tag{9}$$

where $\mathbf{1}_E$ stands for the function identically equal to 1 on E.

Equality (9) links the characteristic function of $S_n(p)$ to the iterated Fourier operator $Q_{\theta,p}(t)^n$. Thus, according to Equality (9), Hypothesis $\mathcal{R}(m)$ requires to study the behavior of the application $t \mapsto Q_{\theta,p}(t)^n$ near 0. A natural assumption to do it is to assume that there exists a Banach space $\mathcal{B}$ which contains the function $\mathbf{1}_E$ and on which $(Q_\theta)_{\theta\in\Theta}$ acts continuously (i.e. $\forall \theta \in \Theta$, $Q_\theta \in \mathcal{L}(\mathcal{B})$) and $(Q_\theta)_{\theta\in\Theta}$ satisfies the following uniform strong ergodicity properties (ERG.1)-(ERG.2):

ERG.1. : *$\{\pi_\theta;\ \theta \in \Theta\}$ is bounded in $\mathcal{B}'$.*

ERG.2. : *The transition kernel $(Q_\theta)_{\theta\in\Theta}$ has a spectral gap on $\mathcal{B}$ (uniformly in θ), that is*

$$\lim_{n\to+\infty} \sup_{\theta\in\Theta} \|Q_\theta^n - \Pi_\theta\|_{\mathcal{B}} = 0,$$

where Π_θ denotes the rank-one projection defined on $\mathcal{B}$ by $\Pi_\theta f := \pi_\theta(f)\mathbf{1}_E$.

More precisely, we use the following equivalent form (ERG.2') of (ERG.2):

ERG.2'. : *There exist $c_0 > 0$ and $0 \leq \kappa_0 < 1$ (independent on $\theta \in \Theta$) such that*

$$\forall \theta \in \Theta, \forall n \in I\!N, \quad \|Q_\theta^n - \Pi_\theta\|_{\mathcal{B}} \leq c_0 \kappa_0^n.$$

Note that under (ERG.2'), for all $\theta \in \Theta$, the spectrum $\sigma(Q_{\theta|\mathcal{B}})$ of Q_θ acting on $\mathcal{B}$ belongs to the set $\{z \in \mathbb{C};\ |z| \leq \kappa_0\} \cup \{1\}$.

Then, to derive the properties of $\mathcal{R}(m)$ from (9), we need some spectral perturbation method to control (uniformly in $(\theta, p) \in \Theta \times \mathcal{P}$) the spectrum of $Q_{\theta,p}(t)$ acting on $\mathcal{B}$ whenever $|t|$ is small enough. The usual method requires the continuity at $t = 0$ of the $\mathcal{L}(\mathcal{B})$-valued function $t \mapsto Q_{\theta,p}(t)$, but this continuity assumption involves too strong hypotheses (see [HP10, §3] for details). An alternative method consists in using the Keller-Liverani theorem [KL99, Liv04] (see also [Bal00, Fer]). Using this method, the regularity of $\lambda_{\theta,p}(\cdot)$, $l_{\theta,p}(\cdot)$ and $r_{\theta,p,n}(\cdot)$ is studied in [HP10] in the case of ρ-mixing Markov chains, V-geometrically ergodic Markov chains and for iterated function systems. More exactly, their results are only established for additive univariate functionals of $(X_n)_{n\in\mathbb{N}^*}$, but the extension to our parametric bivariate case (7) is quite natural. This work has already been done in [HLP] in the case of V-geometrically ergodic Markov chains (in Section 4, we will directly use their results).

By contrast, as already mentioned in Introduction, the method developed in [HP10] is not sufficient to investigate Hypothesis (N-A) in our parametric bivariate case. We can all the same easily state the following implication: thanks to (9), provided that the following condition is imposed on $(\mu_\theta)_{\theta\in\Theta}$:

$$\{\mu_\theta;\ \theta\in\Theta\} \text{ is bounded in } \mathcal{B}' \tag{10}$$

and that for all $t\in\mathbb{R}$, $(\theta,p)\in\Theta\times\mathcal{P}$, the operator $Q_{\theta,p}(t)$ belongs to $\mathcal{L}(\mathcal{B})$, the family $\{S_n(p) := \sum_{k=1}^n \xi_p(X_{k-1},X_k); n\in\mathbb{N}^*,\ p\in\mathcal{P}\}$ fulfills Hypothesis (N-A) whenever the following condition holds:

Hypothesis (N-A)' (Operator-type non-arithmeticity). *For any compact subset $K_0\subset\mathbb{R}^*$, there exists $\rho<1$ such that*

$$\forall n\geq 1,\quad \sup\left\{\|Q_{\theta,p}(t)^n\|_{\mathcal{B}};\ t\in K_0,\ (\theta,p)\in\Theta\times\mathcal{P}\right\} = O(\rho^n).$$

In the next section, we replace this condition by some more practical non-lattice conditions.

3 From non-lattice conditions to (N-A)'

We assume that the general Markovian assumptions of the previous Subsection 2.2 hold true. Furthermore, we also assume that there exists a Banach space $\mathcal{B}$ of complex measurable functions defined on E which contains the function $\mathbf{1}_E$, such that for all $\theta\in\Theta$, $\pi_\theta\in\mathcal{B}'$ and such that for all $t\in\mathbb{R}$, $(\theta,p)\in\Theta\times\mathcal{P}$, the Fourier operators $Q_{\theta,p}(t)$ defined in (8) belong to $\mathcal{L}(\mathcal{B})$.

Let us introduce the following non-lattice condition which will be proved (under some additional conditions) to imply the previous operator-type non-arithmetic condition (N-A)'.

Hypothesis (N-L) (Non-lattice). *There exist no $(\theta_0,p_0)\in\Theta\times\mathcal{P}$, no real $a=a(\theta_0,p_0)$, no closed subgroup $H=c\mathbb{Z}$ with $c=c(\theta_0,p_0)\in\mathbb{R}^*$, no π_{θ_0}-full Q_{θ_0}-absorbing set[3] $A=A(\theta_0,p_0)\in\mathcal{E}$, and finally no measurable bounded function $\alpha=\alpha(\theta_0,p_0):E\to\mathbb{R}$ such that*

$$\forall x\in A,\quad \xi_{p_0}(x,y)+\alpha(y)-\alpha(x)\in a+H\quad Q_{\theta_0}(x,dy)-a.s.. \tag{11}$$

3.1 Intermediate conditions

The link between (N-L) and (N-A)' is based on the three following operator-type properties. The first one concerns a control of the spectral radius of $Q_{\theta,p}(t)$ acting on $\mathcal{B}$ denoted by

[3] A set $A\in\mathcal{E}$ is said to be π_{θ_0}-full if $\pi_{\theta_0}(A)=1$ and Q_{θ_0}-absorbing if $Q_{\theta_0}(a,A)=1$ for all $a\in A$.

$r(Q_{\theta,p}(t)_{|\mathcal{B}})$:

$$\forall t \neq 0,\ \forall(\theta,p) \in \Theta \times \mathcal{P}, \quad r\left(Q_{\theta,p}(t)_{|\mathcal{B}}\right) < 1. \tag{i}$$

The second property consists in assuming that one has for any compact subset $K_0 \subset I\!R^*$

$$r_{K_0} := \sup\left\{r\left(Q_{\theta,p}(t)_{|\mathcal{B}}\right);\ t \in K_0,\ (\theta,p) \in \Theta \times \mathcal{P}\right\} < 1. \tag{ii}$$

Notice that, whenever (ii) holds true, for all $z \in \mathbb{C}$, $|z| > r_{K_0}$ and for all $t \in K_0$, $(\theta,p) \in \Theta\times\mathcal{P}$, the resolvent operator $(z - Q_{\theta,p}(t))^{-1}$ is well-defined in $\mathcal{L}(\mathcal{B})$. Then the last property consists in assuming that there exists $\rho_0 \in [r_{K_0}, 1)$ such that, for all $\rho \in (\rho_0, 1)$,

$$\sup\left\{\|\,(z - Q_{\theta,p}(t))^{-1}\,\|_{\mathcal{B}};\ t \in K_0,\ (\theta,p) \in \Theta \times \mathcal{P},\ |z| = \rho\right\} < +\infty. \tag{iii}$$

Below we study the following implications:

(a) (N-L) $\Rightarrow$ (i) under some conditions (and even better: (N-L) $\Leftrightarrow$ (i) under some more conditions)

(b) (i) $\Rightarrow$ (ii)-(iii) under some conditions;

(c) (ii)-(iii) $\Rightarrow$ (N-A)'.

The main difficulty is the proof of the statement (b). For this part, three methods are proposed in Subsection 3.3. Notice that the operator-type non-arithmetic condition (N-A)' obviously implies Property (i).

3.2 From the non-lattice condition (N-L) to Property (i)

The following lemma is an easy extension of [HP10, §12] to our parametric bivariate case.

Lemma 1. *Assume that the following assumptions hold true:*

1. *For all $\theta \in \Theta$, $\lambda \in \mathbb{C}$ such that $|\lambda| \geq 1$, and for all $f \in \mathcal{B}$, $f \neq 0$, we have*

$$\left[\forall n \geq 1, \quad |\lambda|^n |f| \leq Q_\theta^n |f|\,\right] \quad \Rightarrow \quad \left[\,|\lambda| = 1 \text{ and } |f| = \pi_\theta(|f|) > 0 \quad \pi_\theta - a.s.\right].$$

2. *For all $(\theta,p) \in \Theta \times \mathcal{P}$, $t \in I\!R^*$, there exists $0 \leq \gamma = \gamma(\theta,p,t) < 1$ such that the elements of the spectrum of $Q_{\theta,p}(t)$ acting on $\mathcal{B}$ with modulus greater than γ are isolated eigenvalues of finite multiplicity.*

Assume that (N-L) holds true as well. Then (i) is fulfilled. Moreover Property (i) is equivalent to the following condition: there exist no $t_0 \in I\!R^*$, no $(\theta_0,p_0) \in \Theta\times\mathcal{P}$, no $\lambda = \lambda(\theta_0,p_0,t_0) \in \mathbb{C}$ such that $|\lambda| = 1$, no π_{θ_0}-full Q_{θ_0}-absorbing set $A = A(\theta_0,p_0,t_0) \in \mathcal{E}$ and finally no bounded $w = w(\theta_0,p_0,t_0) \in \mathcal{B}$ such that $|w|_{|A}$ is non-null constant, satisfying

$$\forall x \in A, \quad e^{it_0\xi_{p_0}(x,y)} w(y) = \lambda w(x) \quad Q_{\theta_0}(x,dy) - a.s..$$

The last property of Lemma 1 will not be used later, it is only recalled here for a better understanding.

Remark 2. *In fact, Property (i) is equivalent to (N-L) whenever $e^{i\psi} \in \mathcal{B}$ for all bounded real measurable function ψ on E. Notice that this assumption is obviously fulfilled in the V-geometrically ergodic Markovian model to be studied.*

Assumption 1. of Lemma 1 is always satisfied for strongly ergodic models (cf. (ERG.2)) such that for all $x \in E$, the Dirac distribution δ_x at x belongs to $\mathcal{B}'$. In particular, this assumption is satisfied by the V-geometrically ergodic Markovian model to be studied (other conditions are given in [HP10] to check Assumption 1.).

Assumption 2. is much more difficult to be checked. For now, we only mention that it is equivalent to the following condition: for all $(\theta, p) \in \Theta \times \mathcal{P}$, $t \in I\!R^*$, the essential spectral radius $r_{ess}(Q_{\theta,p}(t)_{|\mathcal{B}})$ of $Q_{\theta,p}(t)$ acting on $\mathcal{B}$ is such that $r_{ess}(Q_{\theta,p}(t)_{|\mathcal{B}}) \le \gamma < 1$. Recall that $Q_{\theta,p}(t)$ is said to be quasi-compact on $\mathcal{B}$ whenever $r_{ess}(Q_{\theta,p}(t)_{|\mathcal{B}}) < r(Q_{\theta,p}(t)_{|\mathcal{B}})$.

3.3 Three methods for Condition (i) to imply (ii)-(iii)

To obtain the implication (i) $\Rightarrow$ (ii)-(iii), we can use one of the following three approaches, in which the sets Θ and $\mathcal{P}$ are assumed to be compact.

- First approach. Using the standard operator perturbation theory, specifically the upper-semi-continuity of the function "spectral radius" (see e.g. [HH01, p 19]), one can prove the following statement:

  Assume that $\|Q_{\theta,p}(t) - Q_{\theta_0,p_0}(t_0)\|_{\mathcal{B}} \to 0$ when $(t, \theta, p) \to (t_0, \theta_0, p_0)$. Then the implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) are true.

However, as already mentioned in Subsection 2.2, the last assumption of continuity of $t \mapsto Q_{\theta,p}(t)$ is too restrictive. That is why we will not apply this approach in this work.

- Second approach. It consists in using the perturbation Keller-Liverani theorem instead of the standard perturbation theory. The proof of the following proposition is not provided in this paper since it is an easy extension of [HP10, lem 12.3].

  Proposition 2. *Assume that there exists some semi normed space $\widetilde{\mathcal{B}}$ such that for all $t \in I\!R$ and $(\theta, p) \in \Theta \times \mathcal{P}$, $Q_{\theta,p}(t)$ belongs to $\mathcal{L}(\widetilde{\mathcal{B}})$ and $\mathcal{B} \hookrightarrow \widetilde{\mathcal{B}}$ (i.e. $\mathcal{B} \subset \widetilde{\mathcal{B}}$ and the identity map is continuous from $\mathcal{B}$ into $\widetilde{\mathcal{B}}$). Furthermore assume that for all $t_0 \in I\!R^*$ and $(\theta_0, p_0) \in \Theta \times \mathcal{P}$, there exists a neighborhood $\widetilde{I_0} \subset I\!R$ of (t_0, θ_0, p_0) such that*

  (C1) there exist $c > 0$, $0 \le \kappa < 1$ and $M > 0$ such that for all $(t, \theta, p) \in \widetilde{I_0}$, $f \in \mathcal{B}$, $n \in I\!N$, one has
  $\|Q_{\theta,p}(t)^n f\|_{\mathcal{B}} \le c\ \kappa^n \|f\|_{\mathcal{B}} + c\ M^n\ \|f\|_{\widetilde{\mathcal{B}}}$.

  (C2) $\|Q_{\theta,p}(t) - Q_{\theta_0,p_0}(t_0)\|_{\mathcal{B},\widetilde{\mathcal{B}}} \to 0$ *when* $(t, \theta, p) \to (t_0, \theta_0, p_0)$.

Then the implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) are true.

This second approach is applied in Subsection 6.4 to some AR(d) processes with $d \geq 2$.

Note that Condition (C2) may be difficult to be checked because of the continuity with respect to θ. However, under the standard dominated model assumption, Condition (C2) can be dodged using the following approach.

- Third approach. It consists in using quasi-compactness and Ascoli-type arguments. For instance, let us give a brief account for the implication (i) $\Rightarrow$ (ii). We assume by absurd that (ii) does not hold and (i) holds true, namely on the one hand there exists a compact subset K_0 of $\mathbb{R}^*$ such that $r_{K_0} := \sup\{r(Q_{\theta,p}(t)_{|\mathcal{B}});\ t \in K_0,\ (\theta,p) \in \Theta \times \mathcal{P}\} \geq 1$, and on the other hand $r_{K_0} \leq 1 < +\infty$. Then there exist some sequences $(t_k)_{k\in\mathbb{N}} \in K_0^{\mathbb{N}}$ and $(\theta_k, p_k)_{k\in\mathbb{N}} \in (\Theta \times \mathcal{P})^{\mathbb{N}}$ such that $\lim r(Q_{\theta_k,p_k}(t_k)) \geq 1$ when $k \to +\infty$. Under the quasi-compactness Assumption 2. of Lemma 1, the previous property implies the existence of $(\lambda_k)_{k\in\mathbb{N}} \in \mathbb{C}^{\mathbb{N}}$ and $(w_k)_{k\in\mathbb{N}} \in \mathcal{B}^{\mathbb{N}}$ such that $Q_{\theta_k,p_k}(t_k)w_k = z_k w_k$ and $|\lambda_k| = r(Q_{\theta_k,p_k}(t_k))$. Finally, from compactness arguments (in particular by using Ascoli theorem), there exist some $\tilde{t} \in K_0$, $(\tilde{\theta}, \tilde{p}) \in \Theta \times \mathcal{P}$, $\tilde{\lambda} \in \mathbb{C}$, $|\tilde{\lambda}| \geq 1$, and $\tilde{w} \in \mathcal{B}$ such that $Q_{\tilde{\theta},\tilde{p}}(\tilde{t})\tilde{w} = \tilde{\lambda}\tilde{w}$, which is in contradiction with (i). Similar arguments can be used to prove (ii) $\Rightarrow$ (iii). In practice, it is easier to use Ascoli theorem when the model is dominated (i.e. $\forall x \in E$, $\forall \theta \in \Theta$, $Q_\theta(x, dy) = q_\theta(x, y)\,\mu(dy)$) with suitable conditions on the function $(q_\theta)_{\theta\in\Theta}$ and on the dominating positive measure μ.

  This approach is detailed in Subsection 4.2 for V-geometrically ergodic Markov chains and then it is applied in Subsection 5.2 to AR(1) processes with Gaussian noise.

3.4 From Properties (ii)-(iii) to the operator-type non-arithmetic condition (N-A)'

Lemma 2. *Assume that Properties (ii)-(iii) hold true. Then (N-A)' is fulfilled.*

Proof of Lemma 2. Let $K_0 \subset \mathbb{R}^*$ be any compact set and let Γ denote the oriented circle defined by $\{z \in \mathbb{C};\ |z| = \rho\}$ where $\rho \in (\rho_0, 1)$. From Von Neumann series, we have for all $t \in K_0$ and $(\theta, p) \in \Theta \times \mathcal{P}$,

$$z \in \mathbb{C},\ |z| = \rho \quad \Rightarrow \quad (z - Q_{\theta,p}(t))^{-1} = \sum_{n=0}^{+\infty} z^{-n-1} Q_{\theta,p}(t)^n$$

and hence, we obtain

$$\forall t \in K_0, \forall (\theta,p) \in \Theta \times \mathcal{P}, \forall n \geq 1, \quad Q_{\theta,p}(t)^n = \frac{1}{2i\pi} \int_\Gamma z^n\ (z - Q_{\theta,p}(t))^{-1}\, dz.$$

Then (N-A)' can easily be derived thanks to (iii). □

3.5 Conclusion

Subsections 3.2, 3.3 and 3.4 give a procedure to derive (N-A)' (and so (N-A)) from the non-lattice condition (N-L). In some cases, we may need some even more simple condition than (N-L) to check (N-A). However, notice that this new condition, denoted by (N-L)', is not equivalent to (N-L).

Assume that the set E is topological and let $\mathcal{E} := \mathcal{B}(E)$ be the associated Borel algebra.

Hypothesis (N-L)'. *For all $p \in \mathcal{P}$, there do not exist $\mathcal{A}_p(\cdot)$ and C_p such that we have for all $(x,y) \in E^2$, $\xi_p(x,y) = \mathcal{A}_p(y) - \mathcal{A}_p(x) + C_p$.*

To connect (N-L)' with (N-L), we need the following hypotheses on both the model and $(\xi_p)_{p\in\mathcal{P}}$.

Hypothesis ($\mathcal{S}$). *There exists a positive measure μ on E satisfying $Supp(\mu) = E$ and such that we have for any $B \in \mathcal{E}$:*

$$[\ \exists\ (\theta,x) \in \Theta \times E,\quad Q_\theta(x,B) = 0\ ] \quad \Longrightarrow \quad [\ \mu(B) = 0\ ].$$

Hypothesis ($\mathcal{C}$). *For all $p \in \mathcal{P}$, the application ξ_p is continuous from E^2 into $\mathbb{R}$.*

Lemma 3. *Assume that the set E is connex and that Assumptions ($\mathcal{S}$) and ($\mathcal{C}$) hold true. If the family $(\xi_p)_{p\in\mathcal{P}}$ fulfills (N-L)', then (N-L) is fulfilled.*

Proof of Lemma 3. Assume that (N-L) is not fulfilled, that is we have (11) with some $(\theta_0, p_0) \in \Theta \times \mathcal{P}$, $a \in \mathbb{R}$, some closed subgroup $H = c\mathbb{Z}$ ($c \in \mathbb{R}^*$), some π_{θ_0}-full Q_{θ_0}-absorbing set $A \in \mathcal{E}$, and finally some bounded measurable function $\alpha : E \to \mathbb{R}$. For the sake of simplicity, let us omit the dependence on (θ_0, p_0). For all $x \in A$, there exists $E_x \in \mathcal{E}$ such that $Q(x, E_x) = 1$ and $\forall y \in E_x$, $\xi(x,y) + \alpha(y) - \alpha(x) \in a + H$. Let $x_0 \in A$. One has

$$\begin{aligned} \forall y \in E_{x_0}, \qquad & \xi(x_0,y) + \alpha(y) - \alpha(x_0) &\in\ & a + H \\ \forall x \in E_{x_0}, \qquad & \xi(x_0,x) + \alpha(x) - \alpha(x_0) &\in\ & a + H. \end{aligned}$$

Thanks to Assumption ($\mathcal{S}$), one has $\mu(E \backslash E_{x_0}) = 0$ and $\mu(E\backslash A) = 0$ (recall that A is Q-absorbing), and hence $\mu(E\backslash\{A \cap E_{x_0}\}) = 0$, $\overline{A \cap E_{x_0}} \supset Supp(\mu) = E$ where $\overline{A \cap E_{x_0}}$ denotes the closure of $A \cap E_{x_0}$. In particular, $A \cap E_{x_0}$ is not empty. Let $x \in A \cap E_{x_0}$, then

$$\forall y \in E_{x_0} \cap E_x, \qquad \xi(x,y) - (\xi(x_0,y) - \xi(x_0,x)) \ \in\ a + H.$$

Let us define $\mathcal{A}(x) := \xi(x_0,x)$ and $f(x,y) := \xi(x,y) + \mathcal{A}(x) - \mathcal{A}(y)$. Then for all $x \in A \cap E_{x_0}$, $f(x, E_{x_0} \cap E_x) \subset a + H$. Then, by continuity arguments and since $E = Supp(\mu) = \overline{E_{x_0} \cap E_x}$, one can easily show that $f(x,E) \subset a + H$. In the same way, $f(A \cap E_{x_0}, E) \subset a + H$, and finally $f(E,E) \subset a + H$. Since $f(E,E)$ is connex and $a + H$ is discrete, f is constant on E^2.
□

Remark 3. *Let μ be a positive measure on E satisfying $Supp(\mu) = E$. Assume that the following dominated model condition holds: for all $\theta \in \Theta$, there exists a non-negative measurable application $q_\theta(\cdot,\cdot)$ on $(E \times E, \mathcal{E} \otimes \mathcal{E})$ such that for all $x \in E$, $B \in \mathcal{E}$, $Q_\theta(x,B) = \int_B q_\theta(x,y)\, d\mu(y)$ and for all $x \in E$ and for μ-almost all $y \in E$, $q_\theta(x,y) > 0$. Then one can show that Assumption ($\mathcal{S}$) holds true.*

4 The case of uniform V-geometrically ergodic Markov chains

In this section, we illustrate the previous results for uniform V-geometrically ergodic Markov chains. From now on, for the sake of simplicity, we consider that $E := \mathbb{R}^d$ (with $d \in \mathbb{N}^*$), equipped with any norm $\|\cdot\|$, and μ_d^{Leb} denotes the Lebesgue-measure on E. Let us assume that Θ is a compact set. We introduce the uniform (in $\theta \in \Theta$) V-geometrically ergodic Markovian model, which satisfies Properties (ERG.1)-(ERG.2) on the weighted-supremum normed space associated with V.

Model $(\mathcal{M})$. *For all $\theta \in \Theta$, there exist both a Q_θ-invariant probability measure denoted by π_θ and an unbounded function $V : E \to [1, +\infty)$ such that*

(VG1) $\sup_{\theta\in\Theta} \pi_\theta(V) < +\infty$,

(VG2) $\lim_{n\to+\infty} \sup\Big\{ |Q_\theta^n f(x) - \pi_\theta(f)|/V(x);\ f : E \to \mathbb{C}$ *measurable*, $|f| \le V,\ x \in E,\ \theta \in \Theta\Big\} = 0.$

Model $(\mathcal{M})$ has already been considered for statistical investigation, see for instance [Fuh06, DY07, HLP]. When θ is fixed and when the Markov chain is irreducible and aperiodic, (VG1) and (VG2) can be checked using the so-called drift-criterion, we refer to [MT93, p 367] for details. Notice that Condition (10) on initial the distribution is equivalent to the following one for Model $(\mathcal{M})$:

$$\sup_{\theta\in\Theta} \mu_\theta(V) < +\infty. \tag{12}$$

In the next Subsections 4.1 and 4.2, we consider a family $(\xi_p)_{p\in\mathcal{P}}$ of measurable functions from E^2 into $\mathbb{R}$, with $\mathcal{P}$ assumed to be a compact set, and we successively study Hypotheses $\mathcal{R}(m)$ and (N-A) for Model $(\mathcal{M})$, before applying these results in Section 5 to M-estimators.

4.1 Study of Hypothesis $\mathcal{R}(m)$

Let us recall the following proposition which has already been proven in [HLP, lem 1].

Proposition 3. *Let us consider a Model $(\mathcal{M})$. Assume on $(\xi_p)_{p\in\mathcal{P}}$ that for all $(\theta, p) \in \Theta\times\mathcal{P}$, ξ_p is centered with respect to the invariant measure family $(\pi_\theta)_{\theta\in\Theta}$ (i.e. for all $(\theta, p) \in \Theta \times \mathcal{P}$, $\mathbb{E}_{\theta,\pi_\theta}[\xi_p(X_0, X_1)] = 0$), and that assume that $(\xi_p)_{p\in\mathcal{P}}$ fulfills the following moment domination condition for some $m \in \mathbb{N}$:*

$$\exists\varepsilon > 0, \quad \sup\left\{ \frac{|\xi_p(x, y)|^{m+\varepsilon}}{V(x) + V(y)};\ (x, y) \in E^2,\ p \in \mathcal{P} \right\} < +\infty. \tag{D_m}$$

Finally assume that the initial distribution family $(\mu_\theta)_{\theta\in\Theta}$ satisfies (12). Then the family $\{S_n(p) := \sum_{k=1}^n \xi_p(X_{k-1}, X_k);\ n \in \mathbb{N}^*,\ p \in \mathcal{P}\}$ satisfies Hypothesis $\mathcal{R}(m)$.

Up to the arbitrarily small real number $\varepsilon > 0$, Condition (D_m) is the expected (with respect to the i.i.d. case) assumption to obtain Hypothesis $\mathcal{R}(m)$ in our model. Indeed, in [DY07], Condition (D_2) is the key assumption to prove the asymptotic normality whereas in [HLP], Condition (D_3) is the key assumption to prove Berry-Esseen bounds. Here one also needs to investigate Hypothesis (N-A).

4.2 Study of Hypothesis (N-A) for dominated Models $(\mathcal{M})$

Further assumptions are required to apply what we called the third approach in Subsection 3.3. Some of them concern the dominated model and the other ones involve the regularity of the applications $(\xi_p)_{p\in\mathcal{P}}$.

Assumption $(\mathcal{S}')$. *For all $\theta \in \Theta$, there exists an application $q_\theta(\cdot,\cdot)$ on E^2 such that*

$$\forall x \in E, \quad Q_\theta(x,dy) = q_\theta(x,y)\,\mu_d^{Leb}(dy).$$

Furthermore for all $x \in E$ and for μ_d^{Leb}-almost all $y \in E$, the application $\theta \mapsto q_\theta(x,y)$ is continuous and there exists $\beta > 0$ such that

- *for all $\theta_0 \in \Theta$, there exists a neighborhood $\mathcal{V}_1 = \mathcal{V}_1(\theta_0)$ of θ_0 such that*

$$\forall x_0 \in E, \quad \lim_{x\to x_0} \sup_{\theta\in\mathcal{V}_1} \int_E V(y)^\beta \, |q_\theta(x,y) - q_\theta(x_0,y)| \, \mu_d^{Leb}(dy) = 0.$$

- *for all $x_0 \in E$ and $\theta_0 \in \Theta$, there exists a neighborhood $\mathcal{V}_2 = \mathcal{V}_2(x_0,\theta_0)$ of θ_0 such that*

$$\int_E V(y)^\beta \sup_{\theta\in\mathcal{V}_2} |q_\theta(x_0,y)| \mu_d^{Leb}(dy) < +\infty.$$

Assumption $(\mathcal{C}')$. *The family $(\xi_p)_{p\in\mathcal{P}}$ satisfies*

- *for all $x \in E$ and for μ_d^{Leb}-almost all $y \in E$, the function $p \mapsto \xi_p(x,y)$ is continuous.*
- *for all $x_0 \in E$ and $p_0 \in \mathcal{P}$, there exist neighborhoods $\mathcal{V}_3 = \mathcal{V}_3(x_0,p_0)$ of x_0 and $\mathcal{V}_4 = \mathcal{V}_4(p_0)$ of p_0, some positive numbers C, υ_1 and υ_2 such that we have*

$$\forall p \in \mathcal{V}_4, \ \forall x \in \mathcal{V}_3, \ \forall y \in E, \quad |\xi_p(x,y) - \xi_p(x_0,y)| \le C\,\|x - x_0\|^{\upsilon_1} V(y)^{\upsilon_2}.$$

Theorem 1. *Let us consider a Model $(\mathcal{M})$, and assume that the preceding assumptions $(\mathcal{S}')$ and $(\mathcal{C}')$ hold true. If the non-lattice condition (N-L) of Section 3 holds true and if the family of initial distributions $(\mu_\theta)_{\theta\in\Theta}$ satisfies (12), then $\{S_n(p) := \sum_{k=1}^n \xi_p(X_{k-1},X_k);\ n \in \mathbb{N}^*,\ p \in \mathcal{P}\}$ satisfies Hypothesis (N-A).*

As discussed in Subsection 2.2, to check Hypothesis (N-A), we need a Banach space $\mathcal{B}$ composed of complex measurable functions defined on E, containing the function $\mathbf{1}_E$, such that for all $\theta \in \Theta$, $\pi_\theta \in \mathcal{B}'$, and such that for all $t \in \mathbb{R}$, $(\theta,p) \in \Theta \times \mathcal{P}$, the Fourier operator $Q_{\theta,p}(t)$ belongs to $\mathcal{L}(\mathcal{B})$. From (VG2), the natural space for this job is the Banach space $\mathcal{B}_1$ composed of measurable functions $f : E \to \mathbb{C}$ such that

$$\|f\|_{\mathcal{B}_1} := \sup_{x\in E} \frac{|f(x)|}{V(x)} < +\infty. \tag{13}$$

Actually, for a technical reason arising in Lemma 5 below, we need to work with another space. Let β be given in Assumption $(\mathcal{S}')$. Without loss of generality, one can suppose that $\beta \in (0,1)$. Then we consider the Banach space $\mathcal{B}_\beta$ composed of measurable functions $f : E \to \mathbb{C}$ such that

$$\|f\|_{\mathcal{B}_\beta} := \sup_{x\in E} \frac{|f(x)|}{V(x)^\beta} < +\infty. \tag{14}$$

Notice that for any Model $(\mathcal{M})$, using the drift-criterion (cf. [MT93]) and Jensen inequality, we can prove that (see [HP10, §10])

$$\lim_{n\to+\infty} \sup_{\theta\in\Theta} \|Q_\theta^n - \Pi_\theta\|_{\mathcal{B}_\beta} = 0. \tag{15}$$

Then, Assumption (ERG.2) of Subsection 2.2 holds true with $\mathcal{B} := \mathcal{B}_\beta$. More precisely, we will use the equivalent form (ERG.2') of (ERG.2): there exist $\widetilde{c}_\beta > 0$ and $0 \le \kappa_\beta < 1$ (independent on $\theta \in \Theta$) such that

$$\forall \theta \in \Theta, \forall n \in I\!N, \quad \|Q_\theta^n - \Pi_\theta\|_{\mathcal{B}_\beta} \le \widetilde{c}_\beta \kappa_\beta^n. \tag{16}$$

Proof of Theorem 1. Let $\hbar$ denote $\hbar := (t, \theta, p) \in I\!R \times \Theta \times \mathcal{P}$ and $Q(\hbar) := Q_{\theta,p}(t)$. First of all, notice that, since $\mathcal{B}_\beta$ is a Banach lattice (i.e. for all $(f,g) \in \mathcal{B}_\beta \times \mathcal{B}_\beta$, $|f| \le |g| \Rightarrow \|f\|_{\mathcal{B}_\beta} \le \|g\|_{\mathcal{B}_\beta}$) and using (15), we can apply [RW97, cor 1.6] to prove that the essential spectral radius of $Q(\hbar)$ satisfies

$$\exists\ 0 \le \kappa < 1 \text{ such that } \forall \hbar_0 \in I\!R^* \times \Theta \times \mathcal{P}, \qquad r_{ess}(Q(\hbar_0)_{|\mathcal{B}_\beta}) \le \kappa. \tag{17}$$

Next, let us sum up the gap from (N-L) to (N-A), specifying their link with all the intermediate conditions introduced in Subsection 3.1:

- thanks to the previous Inequality (17) on the essential spectral radius of $Q(\hbar)$, Assumption 2. of Lemma 1 holds true (Assumption 1. of Lemma 1 also holds true: see the comments after Lemma 1). Thus the conclusions of Lemma 1 are satisfied: (N-L) $\Rightarrow$ (i);
- thanks to Lemma 2: (ii)-(iii) $\Rightarrow$ (N-A)' with $\mathcal{B} := \mathcal{B}_\beta$;
- from Condition (12): (N-A)' $\Rightarrow$ (N-A).

Next, it only remains to prove that (i) $\Rightarrow$ (ii)-(iii). In fact, we show that (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii), using quasi-compactness and Ascoli-type arguments, as announced in the third approach of Subsection 3.3. The proof of (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) involves the two following Lemmas 4 and 5.

Lemma 4 (Doeblin-Fortet Inequality). *For any Model $(\mathcal{M})$, there exist $c_\beta > 0$, $0 \le \kappa_\beta < 1$, such that*

$$\forall \hbar \in I\!R \times \Theta \times \mathcal{P},\ \forall f \in \mathcal{B}_\beta,\ \forall n \in I\!N, \qquad \|Q(\hbar)^n f\|_{\mathcal{B}_\beta} \le c_\beta \kappa_\beta^n \|f\|_{\mathcal{B}_\beta} + c_\beta \|f\|_{\mathcal{B}_1}. \tag{D-F}$$

Proof of Lemma 4. Doeblin-Fortet Inequality (D-F) is a consequence of $\|Q_{\theta,p}(t)^n(f)\|_{\mathcal{B}_\beta} \le \|Q_\theta^n(|f|)\|_{\mathcal{B}_\beta}$ (since $\mathcal{B}_\beta$ is a Banach lattice) and (16) and (VG1). Indeed, for all $f \in \mathcal{B}_\beta$, $|f| \in \mathcal{B}_\beta$, and hence one has $\|Q_\theta^n(|f|) - \pi_\theta(|f|)\|_{\mathcal{B}_\beta} \le \widetilde{c}_\beta \kappa_\beta^n \|f\|_{\mathcal{B}_\beta}$, from which we easily deduce the desired inequality. □

Lemma 5. *Let $(w_k)_{k\in I\!N} \in (\mathcal{B}_\beta)^{I\!N}$ such that $\|w_k\|_{\mathcal{B}_\beta} = 1$ for all $k \geq 1$. If $(w_k)_{k\in I\!N}$ uniformly converges to $\widetilde{w} \equiv 0$ on any compact subset of E, then $\sup_{\theta\in\Theta} \|w_k\|_{\mathcal{B}_1} \to 0$ when $k \to +\infty$.*

Proof of Lemma 5. Let $\widetilde{\varepsilon} > 0$, $\varepsilon := 1-\beta$, and let $K = K_{\widetilde{\varepsilon},\varepsilon}$ be a compact subset of E such that $\sup_{x\in E\setminus K} V(x)^{-\varepsilon} \leq \widetilde{\varepsilon}$. Since $|w_k(x)| \leq \|w_k\|_{\mathcal{B}_\beta} V(x)^\beta = V(x)^\beta$, one has

$$\forall k \in I\!N, \quad \|w_k \mathbf{1}_{|E\setminus K}\|_{\mathcal{B}_1} \leq \|V^\beta \mathbf{1}_{|E\setminus K}\|_{\mathcal{B}_1} = \sup_{x\in E\setminus K} \frac{V(x)^\beta}{V(x)} \leq \widetilde{\varepsilon}. \tag{18}$$

Furthermore $\sup_{x\in K} |w_k(x)|/V(x) \leq \sup_{x\in K} |w_k(x)| \to_k 0$, thus there exists $k_0 \in I\!N$ such that

$$\forall k \geq k_0, \quad \sup_{x\in K} \frac{|w_k(x)|}{V(x)} = \|w_k \mathbf{1}_{|K}\|_{\mathcal{B}_1} \leq \widetilde{\varepsilon}. \tag{19}$$

By combining (18) and (19), $\|w_k\|_{\mathcal{B}_1} = \max\left(\|w_k \mathbf{1}_{|E\setminus K}\|_{\mathcal{B}_1}, \|w_k \mathbf{1}_{|K}\|_{\mathcal{B}_1}\right) \leq \widetilde{\varepsilon}$. □

We are now ready to complete the proof of Theorem 1.

Lemma 6. *We have (i) ⇒ (ii).*

Proof of Lemma 6. We assume by absurd that (ii) does not hold and (i) holds true, namely on the one hand there exists a compact subset K_0 of $I\!R^*$ such that $r_{K_0} := \sup\{r(Q(\hbar)_{|\mathcal{B}_\beta}),\ \hbar \in K_0 \times \Theta \times \mathcal{P}\} \geq 1$, and on the other hand $r_{K_0} \leq 1 < +\infty$. Thus there exists $(\hbar_k)_{k\in I\!N} \in (K_0 \times \Theta \times \mathcal{P})^{I\!N}$ such that $\lim r(Q(\hbar_k)_{|\mathcal{B}_\beta}) = r_{K_0}$ when $k \to +\infty$, and for all $k \geq 0$, $r(Q(\hbar_k)_{|\mathcal{B}_\beta}) > \kappa$, where κ is defined in Inequality (17) on the essential spectral radius of $Q(\hbar)$. Then for all $k \geq 0$, there exists an eigenvalue λ_k such that $|\lambda_k| = r(Q(\hbar_k)_{|\mathcal{B}_\beta})$. Let $w_k \in \mathcal{B}_\beta$, $w_k \neq 0$, $\|w_k\|_\beta = 1$, such that

$$Q(\hbar_k) w_k = \lambda_k w_k. \tag{20}$$

By compacity argument, we can suppose $\lim \hbar_k := \widetilde{\hbar} = (\widetilde{t}, \widetilde{\theta}, \widetilde{p})$ and $\lim \lambda_k := \widetilde{\lambda}$ when $k \to +\infty$, with $\widetilde{\hbar} \in K_0 \times \Theta \times \mathcal{P}$ and $|\widetilde{\lambda}| = r_{K_0} \geq 1$.

a) $(w_k)_k$ converges on E to some $\widetilde{w} \in \mathcal{B}_\beta$: Under the first point of $(\mathcal{S}')$ and the second one of $(\mathcal{C}')$, and using Ascoli theorem, it is easy to see that $(Q(\hbar_k)w_k)_{k\geq k_0}$ is relatively compact in $(\mathcal{C}(K, I\!R), \|\cdot\|_\infty)$ for any compact subset K of E. By diagonal extraction, we can suppose that $(Q(\hbar_k)w_k)_{k\in I\!N}$ converges pointwise on E and uniformly on any compact subset of E, and so does the sequence $(w_k)_{k\in I\!N}$ thanks to Equality (20). Its limit is denoted by $\widetilde{w} \in \mathcal{B}_\beta$.

b) $\widetilde{w} \neq 0$: From Doeblin-Fortet Inequality (D-F), from Equality (20) which implies $Q(\hbar_k)^n w_k = \lambda_k^n w_k$ for all $n \in I\!N^*$, and from $\|w_k\|_{\mathcal{B}_\beta} = 1$, one obtains $|\lambda_k|^n \leq c_\beta \kappa_\beta^n + \|w_k\|_{\mathcal{B}_1}$. Suppose that $\widetilde{w} = 0$. Then $\|w_k\|_{\mathcal{B}_1} \to 0$ when $k \to +\infty$ thanks to Lemma 5. Since $|\lambda_k| \to |\widetilde{\lambda}| = r_{K_0}$ when $k \to +\infty$, one has for all $n \in I\!N$, $r_{K_0}^n \leq c_\beta \kappa_\beta^n$, which is in contradiction with the fact that $\kappa_\beta < 1 \leq r_{K_0}$. Consequently $\widetilde{w} \neq 0$.

c) Conclusion: Let $x_0 \in E$. From Assumption $(\mathcal{S}')$, we have

$$Q(\hbar_k) w_k(x_0) \ := \ \int_E w_k(y) e^{it_k \xi_{p_k}(x_0,y)} q_{\theta_k}(x_0, y)\, \mu_d^{\text{Leb}}(dy).$$

Then, under the second point of $(\mathcal{S}')$ and the first one of $(\mathcal{C}')$, and using Lebesgue dominated convergence theorem, one has $Q(\hbar_k)w_k(x_0) \to_k Q(\widetilde{\hbar})\widetilde{w}(x_0)$. We have just proven that there exist $\widetilde{\lambda} \in \mathbb{C}$, $|\widetilde{\lambda}| = r_{K_0} \geq 1$, a non-null function $\widetilde{w} \in \mathcal{B}_\beta$ and finally a parameter $\widetilde{\hbar} \in K_0 \times \Theta \times \mathcal{P}$ such that $Q(\widetilde{\hbar})\widetilde{w} = \widetilde{\lambda}\widetilde{w}$. This fact implies that $\widetilde{\lambda} \in \sigma(Q(\widetilde{\hbar})_{|\mathcal{B}_\beta})$, which is in contradiction with (i). □

Lemma 7. *We have (ii) ⇒ (iii).*

Proof of Lemma 7. Let $K_0 \subset \mathbb{R}^*$ be compact. From (ii), we have $r_{K_0} := \sup\{r(Q(\hbar)_{|\mathcal{B}_\beta}),\ \hbar \in K_0 \times \Theta \times \mathcal{P}\} < 1$. By absurd, we assume that there exists ρ be such that $\max(r_{K_0}, \kappa_\beta) < \rho < 1$ (where κ_β is defined in (D-F)) and such that $\sup_{|z|=\rho} \sup_{\hbar \in K_0 \times \Theta \times \mathcal{P}} \{\| (z - Q(\hbar))^{-1} \|_{\mathcal{B}_\beta}\} = +\infty$. Thus there exist $(\hbar_k, z_k)_{k \in \mathbb{N}} \in (K_0 \times \Theta \times \mathcal{P})^{\mathbb{N}} \times \mathbb{C}^{\mathbb{N}}$, $|z_k| = \rho$, such that $\alpha_k := \| (z_k - Q(\hbar_k))^{-1} \|_{\mathcal{B}_\beta} \to +\infty$ when $k \to +\infty$, which implies by Banach-Steinhaus theorem that there exists $f \in \mathcal{B}_\beta$ satisfying $\| (z_k - Q(\hbar_k))^{-1} f \|_{\mathcal{B}_\beta} \to +\infty$. Let $w_k := (z_k - Q(\hbar_k))^{-1} f/\alpha_k$ and $\varepsilon_k := f/\alpha_k \in \mathcal{B}_\beta$. Then one has

$$Q(\hbar_k)w_k = z_k w_k - \varepsilon_k. \tag{21}$$

By compacity argument, we can suppose that $\lim_{k \to +\infty} \hbar_k := \widetilde{\hbar} = (\widetilde{t}, \widetilde{\theta}, \widetilde{p})$ and $\lim_{k \to +\infty} z_k := \widetilde{z}$, with $\widetilde{\hbar} \in K_0 \times \Theta \times \mathcal{P}$, and $|\widetilde{z}| = \rho$.

a) $(w_k)_k$ converges on E to some $\widetilde{w} \in \mathcal{B}_\beta$: Again from Ascoli theorem, diagonal extraction and (21), we can suppose that $(w_k)_k$ converges pointwise on E and uniformly on any compact set of E, and we denote its limit by $\widetilde{w} \in \mathcal{B}_\beta$.

b) $\widetilde{w} \neq 0$: From (21), one can easily show

$$\forall n \in \mathbb{N}, \quad z_k^n w_k = Q(\hbar_k)^n w_k + \sum_{i=0}^{n-1} z_k^i Q(\hbar_k)^{n-1-i} \varepsilon_k. \tag{22}$$

From (D-F), one has for all $\hbar_k = (t_k, \theta_k, p_k) \in \mathbb{R} \times \Theta \times \mathcal{P}$ and $n \in \mathbb{N}$: $\|Q(\hbar_k)^n \varepsilon_k\|_{\mathcal{B}_\beta} \leq C_n \|\varepsilon_k\|_{\mathcal{B}_\beta}$ where $C_n := c_\beta(\kappa_\beta^n + b_1)$. Recall that $|z_k| = \rho$. Thus considering again Equality (22) and (D-F), we obtain

$$\rho^n \|w_k\|_{\mathcal{B}_\beta} \leq c_\beta \kappa_\beta^n \|w_k\|_{\mathcal{B}_\beta} + \|w_k\|_{\mathcal{B}_1} + \sum_{i=0}^{n-1} \rho^i C_{n-i-1} \|\varepsilon_k\|_{\mathcal{B}_\beta}.$$

Suppose that $\widetilde{w} = 0$, then $\|w_k\|_{\mathcal{B}_1} \to_k 0$ using Lemma 5. Since $\|w_k\|_{\mathcal{B}_\beta} = 1$ and $\|\varepsilon_k\|_{\mathcal{B}_\beta} = \|f\|_{\mathcal{B}_\beta}/\alpha_k \to_k 0$, one has for all $n \in \mathbb{N}$: $\rho^n \leq c_\beta \kappa_\beta^n$, which is in contradiction with the fact that $\rho > \kappa_\beta$. Thus we have just proven that $\widetilde{w} \neq 0$.

c) Conclusion: Using Lebesgue dominated convergence theorem, one has for all $x \in E$: $Q(\hbar_k)w_k(x) \to_k Q(\widetilde{\hbar})\widetilde{w}(x)$. We have just proven that there exist $\widetilde{z} \in \mathbb{C}$, $|\widetilde{z}| = \rho$, a non-null function $\widetilde{w} \in \mathcal{B}_\beta$ and a parameter $\widetilde{\hbar} \in K_0 \times \Theta \times \mathcal{P}$ such that $Q(\widetilde{\hbar})w = \widetilde{z}\widetilde{w}$. This fact implies that $r(Q(\widetilde{\hbar})_{|\mathcal{B}_\beta}) \geq \rho$, which is in contradiction with the fact that $\rho > r_{K_0}$. Thus we have just proven by absurd that (ii) ⇒ (iii). □

5 M-estimators associated with V-geometrically ergodic Markov chains. Examples

Let $(X_n)_{n\geq 0}$ be a Markov chain with state space $E := I\!R^d$ and transition kernel $(Q_\theta(x,\cdot);\ x \in E)$, where θ is a parameter in some compact set Θ. The probability distribution of X_0 is denoted by μ_θ. As before, the underlying probability measure and the associated expectation are denoted by $\mathbb{P}_{\theta,\mu_\theta}$ and $\mathbb{E}_{\theta,\mu_\theta}$.

Let us introduce the parameter of interest $\alpha = \alpha(\theta) \in \mathcal{A}$ where $\mathcal{A}$ is an open interval of $I\!R$. To define the so-called true value of the parameter of interest $\alpha_0 = \alpha_0(\theta) \in \mathcal{A}$, we introduce the empirical mean functional

$$\forall \alpha \in \mathcal{A},\ \forall n \in I\!N^*, \qquad M_n(\alpha) := \frac{1}{n}\sum_{k=1}^{n} F(\alpha, X_{k-1}, X_k), \tag{23}$$

where $F(\cdot,\cdot,\cdot)$ is a real-valued measurable function on $\mathcal{A}\times E^2$. For instance, $-M_n$ may be the log-likelihood of data $(X_0,\ldots,X_n)$. We define α_0 as follows

$$\forall \theta \in \Theta, \qquad \alpha_0(\theta) := \arg\min_{\alpha\in\mathcal{A}} \lim_{n\to+\infty} \mathbb{E}_{\theta,\mu_\theta}[M_n(\alpha)], \tag{24}$$

and its M-estimator is supposed to be well-defined by

$$\forall n \in I\!N^*, \qquad \hat{\alpha}_n := \arg\min_{\alpha\in\mathcal{A}} M_n(\alpha). \tag{25}$$

Our goal is to provide an asymptotic expansion of $\mathbb{P}_{\theta,\mu_\theta}\left\{\sqrt{n}(\hat{\alpha}_n - \alpha_0)/\sigma(\theta) \leq u\right\}$ uniformly in $\theta \in \Theta$ and $u \in I\!R$, where σ is some suitable (asymptotic) standard deviation. As in the i.i.d. case (see for example [Pfa73]), we assume throughout this section that the following hypotheses on $(\hat{\alpha}_n)_{n\in I\!N^*}$ hold true:

HYP. 1. $\forall n \geq 1,\ (\partial M_n/\partial\alpha)(\hat{\alpha}_n) = 0,$

HYP. 2. $\forall d > 0,\ \sup_{\theta\in\Theta} \mathbb{P}_{\theta,\mu_\theta}\{\, |\hat{\alpha}_n - \alpha_0| \geq d \,\} = o(n^{-\frac{1}{2}})$.

Notice that the uniform consistency property (HYP.2) has already been studied in a Markovian context, see for example [Bil61, Rou65, Rao72, Gän72, DY07].

Throughout the sequel, we assume that $(X_n)_{n\in I\!N}$ belongs to the class of Models $(\mathcal{M})$ (namely $(X_n)_{n\in I\!N}$ is V-geometrically ergodic uniformly in θ) and that the family of initial distributions $(\mu_\theta)_{\theta\in\Theta}$ satisfies (12). In particular, this last condition will be satisfied if $\mu_\theta \equiv \pi_\theta$ (see (VG1)), or if $\mu_\theta \equiv \delta_x$, where δ_x is the Dirac distribution at any $x \in E$. Then, under some further conditions on the model and on the function F, we prove[4] that there exists a polynomial function $A_\theta(\cdot)$ such that

$$\sup_{\theta\in\Theta}\sup_{u\in I\!R}\left|\mathbb{P}_{\theta,\mu_\theta}\left\{\frac{\sqrt{n}}{\sigma(\theta)}(\hat{\alpha}_n - \alpha_0) \leq u\right\} - \mathcal{N}(u) - \eta(u)\, n^{-\frac{1}{2}}\, A_\theta(u)\right| = o(n^{-\frac{1}{2}}). \tag{26}$$

[4] A small part of this work has been announced in [Fer10, note without proof].

Notice that the true value of the parameter of interest (see (24)) can also be defined by

$$\forall \theta \in \Theta,\ \ \forall \alpha \in \mathcal{A},\ \ \alpha \neq \alpha_0, \qquad \mathbb{E}_{\theta,\pi_\theta}[F(\alpha, X_0, X_1)] > \mathbb{E}_{\theta,\pi_\theta}[F(\alpha_0, X_0, X_1)].$$

Asymptotic expansions for M-estimators in the Markovian case have already been studied in several papers. Indeed maximum likelihood estimators are fully studied in [Dah89] and [LRZ03] in the specific case of stationary Gaussian processes. Some M-estimators for general non-stationary models are also studied in [GH94] and [Fuh06], but each author needs some additional Cramér-type hypothesis. Here we only need the much weaker non-arithmeticity condition. Furthermore our moment conditions on F and its derivatives are almost optimal with respect to the i.i.d. case, see the comments after Theorem 2.

5.1 Edgeworth expansion for M-estimators for dominated models $(\mathcal{M})$

In addition to the previous assumptions (namely $(X_n)_{n\in\mathbb{N}}$ belongs the class of Models $(\mathcal{M})$ and $(\hat{\alpha}_n)_{n\in\mathbb{N}^*}$ satisfies (HYP.1) and (HYP.2)), we assume that $(X_n)_{n\in\mathbb{N}}$ is dominated, i.e. that Condition $(\mathcal{S}')$ holds true (see its definition in Subsection 4.2). Furthermore, we assume that for all $x \in E$ and for μ_d^{Leb}-almost all $y \in E$, we have $q_\theta(x, y) > 0$.

Let us introduce the assumptions concerning the real-valued measurable function F involved in (23). Assume that the map $\alpha \mapsto F(\alpha, \cdot, \cdot)$ is 3-time-differentiable on $\mathcal{A}$ and let $F^{(j)} := \partial^j F/\partial\alpha^j$ denote the derivatives for $j = 1, 2, 3$. Assume that $F^{(1)}$, $F^{(2)}$, $F^{(3)}$ satisfy the following moment domination condition (D_3):

$$\exists \varepsilon > 0 \ \text{ such that } \ \forall j = 1, 2, 3, \quad \sup\left\{\frac{|F^{(j)}(\alpha, x, y)|^{3+\varepsilon}}{V(x) + V(y)};\ (x, y) \in E^2,\, \alpha \in \mathcal{A}\right\} < +\infty. \tag{27}$$

We introduce for $j = 1, 2, 3$

$$\forall \theta \in \Theta,\ \ m_j(\theta) := \mathbb{E}_{\theta,\pi_\theta}\left[F^{(j)}(\alpha_0, X_0, X_1)\right]\ ,\ \ \sigma_j(\theta)^2 := \lim_{n\to+\infty} \mathbb{E}_{\theta,\pi_\theta}\left[n\left(M_n^{(j)}(\alpha_0) - m_j(\theta)\right)^2\right],$$

where $M_n(\cdot)$ is given in (23) and $M_n^{(j)} := \partial^j M_n/\partial\alpha^j$, and where π_θ is the Q_θ-invariant probability measure given in $(\mathcal{M})$. Then, from (27) and using Proposition 1 and Proposition 3, the functions $\sigma_j(\cdot)$ for $j = 1, 2, 3$ are well-defined and bounded in $\theta \in \Theta$.

We consider the following additional assumptions:

C.1. $m_1 \equiv 0$ *and* $\inf_{\theta\in\Theta} m_2(\theta) > 0$.

C.2. $\inf_{\theta\in\Theta} \sigma_j(\theta) > 0$ *for* $j = 1, 2$.

C.3. *There exists a measurable function* $W : E \to [0, +\infty)$ *of the type* $W = C\ V^\eta$ *for some* $\eta \in (0, 1/2)$ *and* $C > 0$ *such that*

$$\forall (\alpha, \alpha') \in \mathcal{A}^2,\ \forall (x, y) \in E^2,\ \ |F^{(3)}(\alpha', x, y) - F^{(3)}(\alpha, x, y)| \le |\alpha' - \alpha|\,(W(x) + W(y)).$$

Let us introduce some assumptions similar to $(\mathcal{C})$-$(\mathcal{C}')$ (see definitions in Subsections 3.5 and 4.2) concerning the regularity of $(F^{(j)})_{j=1,2,3}$. The function F is supposed to satisfy

C.4. *For all $j = 1, 2, 3$ and $\alpha \in \mathcal{A}$, $F^{(j)}(\alpha, \cdot, \cdot)$ is continuous from E^2 into $I\!R$.*

C. 5. *For all $x_0 \in E$ and $\alpha \in \mathcal{A}$, there exist neighborhoods $\mathcal{V}_3 = \mathcal{V}_3(x_0, \alpha)$ of x_0 and $\mathcal{V}_4 = \mathcal{V}_4(\alpha)$ of α, positive real numbers C, υ_1 and υ_2 such that for all $\alpha' \in \mathcal{V}_4$, $x \in \mathcal{V}_3$ and $y \in E$:*

$$\forall j = 1, 2, 3, \qquad |F^{(j)}(\alpha', x, y) - F^{(j)}(\alpha', x_0, y)| \leq C\,\|x - x_0\|^{\upsilon_1} V(y)^{\upsilon_2}.$$

Theorem 2. *Assume that all the preceding assumptions hold true, that Condition (N-L)' (see definition page 12) is satisfied by the following functions*

(a) $\forall \alpha \in \mathcal{A}$, $\forall j = 1, 2$, $F^{(j)}(\alpha, x, y)$

(b) $\forall \alpha \in \mathcal{A}$, $\forall \upsilon \in I\!R$, $F^{(1)}(\alpha, x, y) + \upsilon F^{(2)}(\alpha, x, y) + (\upsilon^2/2)\ F^{(3)}(\alpha, x, y)$

and that the initial probability measure satisfies (12), namely $\sup_{\theta\in\Theta} \mu_\theta(V) < +\infty$. *Then for* $j = 1, 2, 3,$

$$\forall \theta \in \Theta, \quad m_j(\theta) = \lim_{n\to+\infty} \mathbb{E}_{\theta,\mu_\theta}[M_n^{(j)}(\alpha_0)] \quad , \quad \sigma_j(\theta)^2 = \lim_{n\to+\infty} \mathbb{E}_{\theta,\mu_\theta}\left[n\left(M_n^{(j)}(\alpha_0) - m_j(\theta)\right)^2\right]$$

and there exists a polynomial function denoted by A_θ such that $(\hat{\alpha}_n)_{n\in I\!N^*}$ satisfies Expansion (26) with $\sigma := \sigma_1/m_2$. Furthermore, the coefficients of A_θ are bounded, and

$$A_\theta(u) := \left[-\frac{1}{6}\frac{m_{3,1}(\theta)^3}{\sigma_1(\theta)^3} + \frac{b_1(\theta)}{\sigma_1(\theta)}\right] + \left[\frac{1}{6}\frac{m_{3,1}(\theta)^3}{\sigma_1(\theta)^3} - \frac{\sigma_{12}(\theta)}{\sigma_1(\theta)m_2(\theta)} + \frac{\sigma_1(\theta)}{2m_2(\theta)^2}m_3(\theta)\right]u^2,$$

where

$$\begin{cases} b_1(\theta) & := & \lim_{n\to+\infty} \mathbb{E}_{\theta,\mu_\theta}[n\, M_n^{(1)}(\alpha_0)\,] \\ \sigma_{12}(\theta) & := & \lim_{n\to+\infty} \mathbb{E}_{\theta,\pi_\theta}\left[n\, M_n^{(1)}(\alpha_0)\left(M_n^{(2)}(\alpha_0) - m_2(\theta)\right)\right] \\ & = & \lim_{n\to+\infty} \mathbb{E}_{\theta,\mu_\theta}\left[n\, M_n^{(1)}(\alpha_0)\left(M_n^{(2)}(\alpha_0) - m_2(\theta)\right)\right] \\ m_{3,1}(\theta)^3 & := & \lim_{n\to+\infty} \mathbb{E}_{\theta,\pi_\theta}\left[n^2\left(M_n^{(1)}(\alpha_0)\right)^3\right] = \lim \mathbb{E}_{\theta,\mu_\theta}\left[n^2\left(M_n^{(1)}(\alpha_0)\right)^3\right] - 3\sigma_1^2(\theta)b_1(\theta). \end{cases}$$

When comparing with Pfanzagl results [Pfa73] in the i.i.d. case, Expansion (26) proven in Theorem 2 is a natural substitute of the i.i.d. expansion, with an additional term due to the asymptotic bias (namely $b_1(\cdot)$). To the best of our knowledge, Theorem 2 is the most precise statement concerning the first-order Edgeworth expansion of real-valued M-estimators associated with V-geometrically ergodic Markov chains: in fact, the dominated model condition $(\mathcal{S}')$ on the model is classical in Markovian statistics, the condition (D_3) on the derivatives of F is the expected one (up to the real number $\varepsilon > 0$), Conditions (C.1)-(C.5) are the Markovian substitutes of the so-called regularity conditions of the i.i.d. case and finally the non-lattice-type conditions (a)-(b) in Theorem 2 are quite general and easy to check.

The proof of Theorem 2 is postponed to Section 6.

An application to autoregressive models is given in the next subsection.

5.2 A simple illustration of Theorem 2: a Gaussian linear model

Let us consider a Gaussian linear process $(X_k)_{k\geq 0}$ defined by

$$X_0 := 0 \qquad \forall k \in \mathbb{N}^*, \quad X_k := \theta X_{k-1} + Z_k \tag{28}$$

where

- θ is a parameter which belongs to a compact set $\Theta \subset (-1, 1)$
- $(Z_k)_{k\in\mathbb{N}^*}$ are i.i.d. real valued r.v. whose common distribution is supposed to be the Gaussian distribution $\mathcal{N}(m, \sigma^2)$ with probability density f_Z with respect to μ^{Leb}.

This simple model allows us to easily illustrate the performance of Theorem 2. For the sake of simplicity, we have considered a Gaussian noise and the Dirac distribution δ_0, but another probability density $f_Z > 0$ and another initial distribution μ_θ could be chosen.

The sequence $(X_k)_{k\geq 0}$ is a Markov chain with state space $E := \mathbb{R}$. Its transition kernel Q_θ is given for all Borel set $B \in \mathcal{B}(\mathbb{R})$ by

$$Q_\theta(x, B) = \int_{\mathbb{R}} \mathbf{1}_B(\theta x + z)\, f_Z(z)\, dz.$$

First of all, note that

$$\forall j > k, \quad X_j = \theta^{j-k} X_k + \sum_{l=1}^{j-k} \theta^{j-k-l} Z_{k+l}. \tag{29}$$

Then for all $k \geq 1$, the r.v. X_k has distribution $\mathcal{N}\big(m(1-\theta^k)/(1-\theta), \sigma^2(1-\theta^{2k})/(1-\theta^2)\big)$, and the Markov chain $(X_k)_{k\geq 0}$ converges in distribution to $\pi_\theta := \mathcal{N}\big(m/(1-\theta), \sigma^2/(1-\theta^2)\big)$, so that

$$\mathbb{E}_{\theta,\pi_\theta}[X_0] = \frac{m}{1-\theta} \quad \text{and} \quad \mathbb{E}_{\theta,\pi_\theta}\left[X_0^2\right] = \frac{\sigma^2}{1-\theta^2} + \frac{m^2}{(1-\theta)^2}. \tag{30}$$

For any $\gamma > 8$, let us define the following function V

$$\forall x \in \mathbb{R}, \quad V(x) := 1 + |x|^\gamma. \tag{31}$$

Since θ belongs to a compact set $\Theta \subset (-1, 1)$, $\sup_{\theta\in\Theta} \mathbb{E}_{\theta,\delta_0}[|Z_1|^\gamma] < +\infty$ and $f_Z > 0$, it is easily checked that $(X_k)_{k\geq 0}$ is μ^{Leb}-irreducible, aperiodic and fulfills the drift-criterion [MT93] uniformly in $\theta \in \Theta$. Therefore, the Markov chain $(X_k)_{k\geq 0}$ is an instance of Model $(\mathcal{M})$ with this function V.

Assume that θ is unknown and has to be estimated ($\alpha_0(\theta) \equiv \theta$ here). The parameters m and $\sigma > 0$ are supposed to be known. The maximum likelihood estimator (MLE) $(\hat{\theta}_n)_{n\in\mathbb{N}^*}$ involves the following function F

$$\forall (x, y) \in \mathbb{R}^2,\ \forall \alpha \in \Theta, \quad F(\alpha, x, y) := -\ln f_Z(y - \alpha\, x),$$

and the empirical mean functional (cf. (23))

$$\forall n \in \mathbb{N}^*,\ \forall \alpha \in \Theta, \quad M_n(\alpha) := -\frac{1}{n}\sum_{k=1}^{n} \ln f_Z(X_k - \alpha X_{k-1}). \tag{32}$$

Let us check that Theorem 2 can be applied to this Gaussian linear model. First we easily obtain $M_n^{(1)}(\hat{\theta}_n) = 0$ (that is (HYP.1)) from the definition of the estimator, and we obtain as well the following closed-form expression for the MLE

$$\forall n \in I\!N^*, \quad \hat{\theta}_n = \frac{\sum_{k=1}^n (X_{k-1}X_k - mX_{k-1})}{\sum_{k=1}^n X_{k-1}^2}. \tag{33}$$

Then uniform consistency Property (HYP.2) holds true. Indeed, write

$$\hat{\theta}_n - \theta = \frac{\Delta_{1n} - \theta\Delta_{2n} - m\Delta_{3n}}{\Delta_{2n} + \mathbb{E}_{\theta,\pi_\theta}[X_0^2]},$$

where

- $\Delta_{1n} := (1/n)\sum_{k=1}^n \big(X_{k-1}X_k - \theta\mathbb{E}_{\theta,\pi_\theta}[X_0^2] - m\mathbb{E}_{\theta,\pi_\theta}[X_0]\big)$
- $\Delta_{2n} := (1/n)\sum_{k=1}^n \big(X_{k-1}^2 - \mathbb{E}_{\theta,\pi_\theta}[X_0^2]\big)$
- $\Delta_{3n} := (1/n)\sum_{k=1}^n (X_{k-1} - \mathbb{E}_{\theta,\pi_\theta}[X_0])$.

Since the families $(\xi_p)_{p\in\mathcal{P}}$ fulfill Condition (D_4) when $\xi_p(x,y) := xy$, $\xi_p(x,y) := x^2$ or $\xi_p(x,y) := x$ (because $\gamma > 8$), we obtain from Property (5)

$$\forall i = 1,2,3, \quad \lim_{n\to+\infty} \sup_{\theta\in\Theta} \mathbb{E}_{\theta,\delta_0}\left[\frac{\Delta_{in}^4}{n^2}\right] < +\infty.$$

One concludes using Markov inequality and since $\inf_{\theta\in\Theta}\mathbb{E}_{\theta,\pi_\theta}[X_0^2] > 0$. Next, the moment domination conditions (27) are obviously fulfilled (since $\gamma > 6$). Concerning Conditions (C.1) and (C.2), we obtain using the invariant probability $(\pi_\theta)_{\theta\in\Theta}$:

- $m_1(\theta) = \mathbb{E}_{\theta,\pi_\theta}\left[X_0 f_Z^{(1)}(X_1 - \theta X_0)/f_Z(X_1 - \theta X_0)\right] = \mathbb{E}_{\theta,\pi_\theta}[X_0]\ \mathbb{E}_{\theta,\pi_\theta}\left[f_Z^{(1)}(Z_1)/f_Z(Z_1)\right]$,

  $m_1(\theta) \equiv 0$, and $m_2(\theta) = \mathbb{E}_{\theta,\pi_\theta}\big[X_0^2\big]/\sigma^2$, so that $\inf_{\theta\in\Theta} m_2(\theta) > 0$.
- $\sigma_1(\theta)^2 = \lim_n \mathbb{E}_{\theta,\pi_\theta}\big[(1/n)\big(\sum_{k=1}^n X_{k-1}(Z_k - m)/\sigma^2\big)^2\big] = \lim_n \mathbb{E}_{\theta,\pi_\theta}\big[(1/n)\sum_{k=1}^n \big(X_{k-1}(Z_k - m)/\sigma\big)^2\big]/\sigma^2$. Thus $\sigma_1(\theta) = \sqrt{m_2(\theta)}$ satisfies $\inf_{\theta\in\Theta}\sigma_1(\theta) > 0$.

  $\sigma_2(\theta)^2 = \lim_{n\to+\infty} \mathbb{E}_{\theta,\pi_\theta}\left[(1/n)\big(\sum_{k=1}^n X_{k-1}^2/\sigma^2 - n\, m_2(\theta)\big)^2\right]$ and after some tedious computations using (29), we obtain $\inf_{\theta\in\Theta}\sigma_2(\theta) > 0$.

Conditions (C.3), (C.4) and (C.5) are obviously fulfilled as well. Finally Assumption $(\mathcal{S}')$ on the transition kernels $(Q_\theta)_{\theta\in\Theta}$ with $q_\theta(x,\cdot) := f_Z(\cdot - \theta x)$ is easily checked.

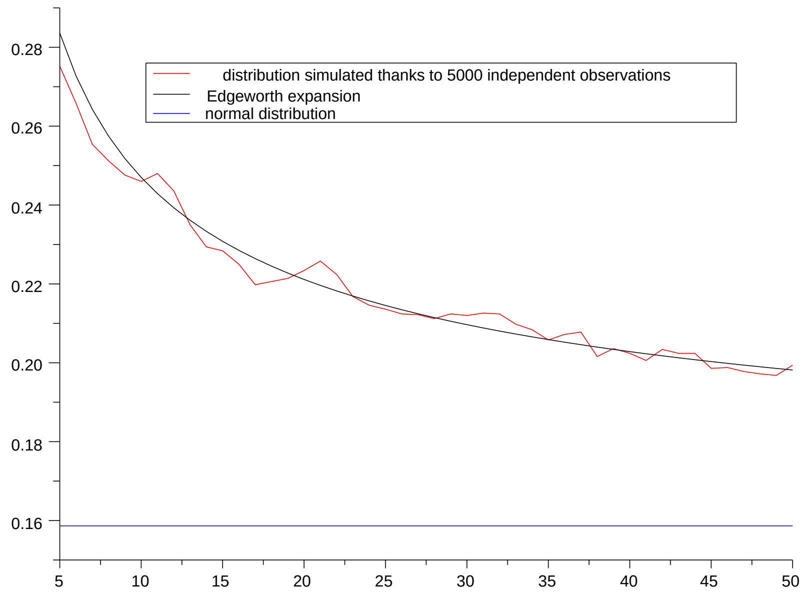
Thus, we apply Theorem 2 to our Gaussian linear model to derive that

$$\sup_{u\in I\!R}\sup_{\theta\in\Theta}\left|\mathbb{P}_{\theta,\delta_0}\left\{\frac{\sqrt{n}}{\sigma(\theta)}(\hat{\theta}_n - \theta) \le u\right\} - \mathcal{N}(u) - \eta(u)\ n^{-\frac{1}{2}}\ A_\theta(u)\right| = o(n^{-\frac{1}{2}}).$$

To illustrate our results, fix $\theta := 1/2$ for example and let us simulate by Monte Carlo methods the behavior of the MLE. Then we can compare the empirical distribution function at u of $\sqrt{n}\,(\hat{\theta}_n - 1/2)/\sigma(1/2)$ with the approximations $\mathcal{N}(u)$ and $\mathcal{N}(u) + \eta(u)A_{1/2}(u)/\sqrt{n}$ of the distribution function $\mathbb{P}_{1/2,\delta_0}\{\sqrt{n}\,(\hat{\theta}_n - 1/2)/\sigma(1/2) \leq u\}$, where the asymptotic standard deviation $\sigma(1/2)$ and the polynomial function $A_{1/2}(u)$ are defined by

$$\sigma\Big(\frac{1}{2}\Big) = \frac{1}{\sqrt{\frac{4}{3} + 4\frac{m^2}{\sigma^2}}}, \qquad A_{\frac{1}{2}}(u) = c(1 + 5u^2) \quad \text{with } c := \frac{8}{18}\frac{\frac{2}{3} + 6\frac{m^2}{\sigma^2}}{\Big(\frac{4}{3} + 4\frac{m^2}{\sigma^2}\Big)^{\frac{3}{2}}}.$$

The performance of this Edgeworth expansion is illustrated in the case where $m = 0$ and $\sigma = 1$. Using a Scilab program, 5000 independent samples of the estimators $(\hat{\theta}_n)_{n=1,..,50}$ have been obtained. For $u := -1$ and $u := 1$, we represent the empirical distribution function of $\sqrt{n}\,(\hat{\theta}_n - 1/2)/\sigma(1/2)$, the Gaussian approximation and the first-order Edgeworth one on the same graphs (see page 24). As expected, observe that the quality of the normal and the first-order Edgeworth approximations $\mathcal{N}(u)$ and $\mathcal{N}(u) + \eta(u)A_{1/2}(u)/\sqrt{n}$ of $\mathbb{P}_{1/2,\delta_0}\{\sqrt{n}\,(\hat{\theta}_n - 1/2)/\sigma(1/2) \leq u\}$ increases when n grows, but the approximation by the first-order expansion is quickly the best one. The first-order Edgeworth expansion greatly improves the rate of convergence in the approximation of the distribution function of the estimator, in comparison with the Berry-Esseen type results.



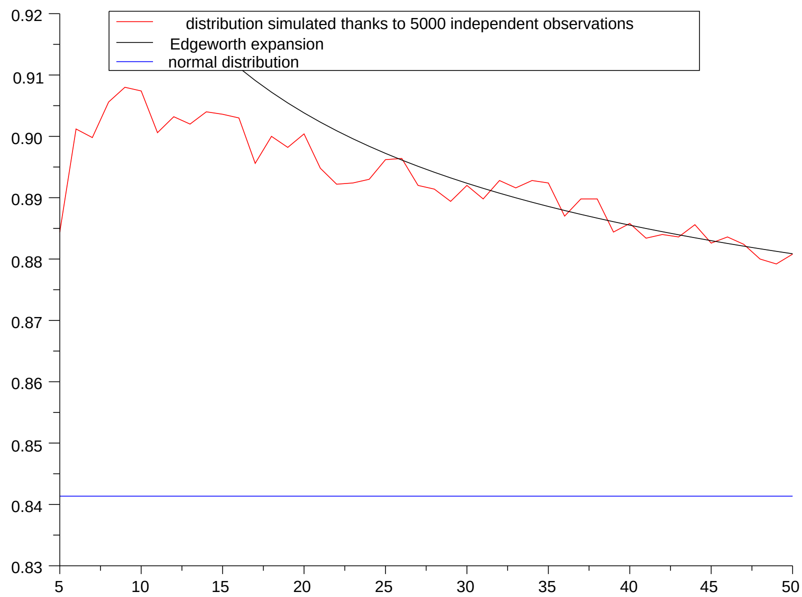


Figure 1: The Gaussian linear example of §5.2: for $u := -1$ and $u := 1$, graphs of
• the empirical distribution of $\sqrt{n}(\hat{\theta}_n - 1/2)/\sigma(1/2)$, i.e. the estimation by Monte-Carlo methods of $n \mapsto \mathbb{P}_{1/2,\delta_0}\left\{\sqrt{n}(\hat{\theta}_n - 1/2)/\sigma(1/2) \leq u\right\}$,
• the Edgeworth expansion $n \mapsto \mathcal{N}(u) + \eta(u)A_{1/2}(u)/\sqrt{n}$
• and the Gaussian approximation $n \mapsto \mathcal{N}(u)$.

6 Pfanzagl method to prove Theorem 2

In Subsections 6.1 and 6.2, we adapt Pfanzagl results to some general setting. More specifically, some probabilistic Edgeworth expansions are explicitly required in Subsection 6.1, whereas the general Assumptions $\mathcal{R}(m)$ and (N-A) are involved in Subsection 6.2. Thanks to this work, Theorem 2 is easily proved in Subsection 6.3.

In Subsections 6.1 and 6.2, we denote by $(\Omega, \mathcal{F}, \{\mathbb{P}_\theta; \theta \in \Theta\})$ a general statistical model, where Θ is some parameter space (not necessarily compact in these subsections). The underlying expectation is denoted by $\mathbb{E}_\theta$. We assume that the following general statistical assumptions hold true: let $(M_n(\alpha))_{n\in\mathbb{N}^*}$ be any general sequence of real observations where $\alpha \equiv \alpha(\theta) \in \mathcal{A}$ is the parameter of interest and $\mathcal{A}$ is an open interval on the real line. Assume that for all $n \geq 1$, the map $\alpha \mapsto M_n(\alpha)$ is 3-time-differentiable on $\mathcal{A}$ and that the derivatives define r.v. on $(\Omega, \mathcal{F})$. We denote them by $M_n^{(j)} := \partial^j M_n / \partial \alpha^j$ for $j = 1, 2, 3$. We consider some $\alpha_0 \in \mathcal{A}$ and assume that the $\mathcal{A}$-valued r.v. $\hat{\alpha}_n$ is specified by (HYP.1) and fulfills the uniform consistency Property (HYP.2) that we recall below:

HYP. 1. $\forall n \geq 1,\ (\partial M_n/\partial\alpha)\,(\hat{\alpha}_n) = 0,$

HYP. 2. $\forall d > 0,\ \sup_{\theta\in\Theta} \mathbb{P}_\theta\big\{\, |\hat{\alpha}_n - \alpha_0| \geq d \,\big\} = o(n^{-1/2})$.

Note that, in Subsections 6.1 and 6.2, $(M_n(\alpha))_{n\in\mathbb{N}^*}$ is not necessarily associated with a function F as in (23).

6.1 The revisited Pfanzagl method

We appeal to the following conditions:

A. 1. *For all $n \geq 1$, there exists a positive r.v. W_n, independent on α, such that*

$$\forall j = 2, 3, \quad \forall (\alpha, \alpha') \in \mathcal{A}^2, \qquad |M_n^{(j)}(\alpha') - M_n^{(j)}(\alpha)| \leq |\alpha' - \alpha|\, W_n.$$

Furthermore there exists $l \colon \Theta \to (0, +\infty)$ bounded on Θ such that $\sup_{\theta\in\Theta} \mathbb{P}_\theta\big\{W_n \geq l(\theta)\big\} = o(n^{-1/2})$.

A. 2. *For $j = 1, 2$, there exist $\sigma_j(\cdot) > 0$ satisfying* $\sup_{\theta\in\Theta} \sigma_j(\theta) < +\infty$, $\inf_{\theta\in\Theta} \sigma_1(\theta) > 0$, *$m_2(\cdot)$ satisfying* $\inf_{\theta\in\Theta} m_2(\theta) > 0$, *and polynomial functions denoted by $B_\theta(\cdot)$ and $C_\theta(\cdot)$, such that*

$$\sup_{\theta\in\Theta} \sup_{u\in\mathbb{R}} \left| \mathbb{P}_\theta \left\{ \frac{\sqrt{n}}{\sigma_1(\theta)}\, M_n^{(1)}(\alpha_0) \leq u \right\} - \mathcal{N}(u) - \eta(u) n^{-\frac{1}{2}} B_\theta(u) \right| = o(n^{-\frac{1}{2}}),$$

$$\sup_{\theta\in\Theta} \sup_{u\in\mathbb{R}} \left| \mathbb{P}_\theta \left\{ \frac{\sqrt{n}}{\sigma_2(\theta)} \big(M_n^{(2)}(\alpha_0) - m_2(\theta)\big) \leq u \right\} - \mathcal{N}(u) - \eta(u) n^{-\frac{1}{2}} C_\theta(u) \right| = o(n^{-\frac{1}{2}}).$$

Furthermore the coefficients of $B_\theta(\cdot)$ and $C_\theta(\cdot)$ are assumed to be bounded.

Let us define $\sigma(\theta) := \sigma_1(\theta)/m_2(\theta)$. Notice that $\sigma(\cdot)$ satisfies $\sup_{\theta\in\Theta}\sigma(\theta) < +\infty$.

A.3. *For all $n \geq 1$, $u \in I\!R$ such that $|u| \leq 2\sqrt{\ln n}$, there exist $\sigma_{n,u}(\cdot)^2 > 0$, $m_3(\cdot)$ bounded on Θ, $D_\theta(\cdot)$ and $E_\theta(\cdot)$ some polynomial functions such that*

$$\sup_{\theta\in\Theta}\sup_{|u|\leq 2\sqrt{\ln n}}\left|\sigma_{n,u}(\theta)^{-1} - \left(\sigma_1(\theta)^{-1} + D_\theta(u)\,n^{-\frac{1}{2}}\right)\right| = o(n^{-\frac{1}{2}}),$$

$$\sup_{\theta\in\Theta}\sup_{v\in I\!R}\sup_{|u|\leq 2\sqrt{\ln n}}\left|\mathbb{P}_\theta\left\{\frac{\sqrt{n}}{\sigma_{n,u}(\theta)}\widetilde{M_n}(\theta,u) \leq v\right\} - \mathcal{N}(v) - \eta(v)E_\theta(v)n^{-\frac{1}{2}}\right| = o(n^{-\frac{1}{2}}),$$

where $\widetilde{M_n}(\theta,u)$ denotes

$$\widetilde{M_n}(\theta,u) := M_n^{(1)}(\alpha_0) + \frac{\sigma(\theta)}{\sqrt{n}}u\Big(M_n^{(2)}(\alpha_0) - m_2(\theta)\Big) + \left(\frac{\sigma(\theta)}{\sqrt{n}}\right)^2\frac{u^2}{2}\Big(M_n^{(3)}(\alpha_0) - m_3(\theta)\Big). \quad (34)$$

Furthermore, the coefficients of $D_\theta(\cdot)$ and $E_\theta(\cdot)$ are assumed to be bounded.

Theorem 3. *Under Conditions (A.1), (A.2) and (A.3), there exists a polynomial function A_θ such that one has with $\sigma := \sigma_1/m_2$*

$$\sup_{\theta\in\Theta}\sup_{u\in I\!R}\left|\mathbb{P}_\theta\left\{\frac{\sqrt{n}}{\sigma(\theta)}(\hat{\alpha}_n - \alpha_0) \leq u\right\} - \mathcal{N}(u) - \eta(u)n^{-\frac{1}{2}}A_\theta(u)\right| = o(n^{-\frac{1}{2}}). \quad (35)$$

Furthermore

$$\forall\theta\in\Theta,\ \forall u\in I\!R, \qquad A_\theta(u) := D_\theta(u)\ \sigma_1(\theta)\ u + \frac{\sigma(\theta)^2}{2\sigma_1(\theta)}m_3(\theta)u^2 - E_\theta(-u). \quad (36)$$

The proof of Theorem 3 is postponed to Appendix A. It consists in adapting the Pfanzagl method [Pfa73] introduced for i.i.d. observations. Just mention that the Pfanzagl method is not exactly the one developed in Appendix A, but for convenience this discussion is omitted.

6.2 An alternative statement using Hypotheses $\mathcal{R}(m)$ and (N-A)

Below we appeal to the following assumptions involving Hypotheses $\mathcal{R}(m)$ and (N-A) of Subsection 2.1:

B.1. *For all $n \geq 1$, there exists a positive r.v. W_n, independent on α, such that*

$$\forall j = 2,3,\quad \forall(\alpha,\alpha')\in\mathcal{A}^2, \qquad |M_n^{(j)}(\alpha') - M_n^{(j)}(\alpha)| \leq |\alpha' - \alpha|\,W_n.$$

Furthermore there exists $\widetilde{l}$: $\Theta \to (-1,+\infty)$ bounded on Θ such that $\big\{n(W_n - \widetilde{l}(\theta));\ n \geq 1,\ \theta\in\Theta\big\}$ fulfills Hypothesis $\mathcal{R}(2)$.

B.2. *The family $\big\{n\,M_n^{(1)}(\alpha_0);\ n\geq 1,\ \theta\in\Theta\big\}$ fulfills Hypotheses $\mathcal{R}(3)$ and (N-A). Furthermore there exists $m_2(\cdot)$ on Θ satisfying $\inf_{\theta\in\Theta} m_2(\theta) > 0$ such that $\big\{n\left(M_n^{(2)}(\alpha_0) - m_2(\theta)\right);\ n \geq 1,\ \theta\in\Theta\big\}$ fulfills both Hypotheses $\mathcal{R}(3)$ and (N-A).*

Thanks to the last Assumption (B.2) and Proposition 1, we can define the asymptotic variances

$$\sigma_1(\theta)^2 := \lim_{n\to+\infty} \mathbb{E}_\theta\left[n\left(M_n^{(1)}(\alpha_0)\right)^2\right], \quad \sigma_2(\theta)^2 := \lim_{n\to+\infty} \mathbb{E}_\theta\left[n\left(M_n^{(2)}(\alpha_0) - m_2(\theta)\right)^2\right].$$

Furthermore we assume that the following conditions on these asymptotic variances hold true

B. 3. $\inf_{\theta\in\Theta} \sigma_1(\theta) > 0$,

B. 4. $\inf_{\theta\in\Theta} \sigma_2(\theta) > 0$.

The following additional conditions are also required:

B. 5. *There exists $m_3(\cdot)$, bounded on Θ, such that $\{n\left(M_n^{(3)}(\alpha_0) - m_3(\theta)\right);\, n \geq 1,\, \theta \in \Theta\}$ fulfills Hypothesis $\mathcal{R}(2)$, and $\{n\,\widetilde{M_n}(\theta,u);\, n \geq 1,\, \theta \in \Theta,\, |u| \leq 2\sqrt{\ln n}\}$ fulfills both Hypotheses $\mathcal{R}(3)$ and (N-A) as well, where $\widetilde{M_n}(\theta,u)$ is defined by (34).*

Theorem 4. *Under Assumptions (B.1) to (B.5), there exists a polynomial function A_θ independent on n such that one has (35) with $\sigma := \sigma_1/m_2$. The polynomial function A_θ is of the type $A_\theta(u) = a_1(\theta) + a_2(\theta)u^2$ where, for $i = 1, 2$, $\sup_{\theta\in\Theta} |a_i(\theta)| < +\infty$. Furthermore if we suppose that the additional moment condition holds true:*

$$\forall j = 1,2,3, \quad \forall\theta \in \Theta, \quad \forall n \in I\!N^*, \quad \mathbb{E}_\theta\left[\left|M_n^{(j)}(\alpha_0)\right|^3\right] < +\infty,$$

then one has more precisely

$$a_1(\theta) := -\frac{1}{6}\frac{m_{3,1}(\theta)^3}{\sigma_1(\theta)^3} + \frac{b_1(\theta)}{\sigma_1(\theta)}, \quad a_2(\theta) := \frac{1}{6}\frac{m_{3,1}(\theta)^3}{\sigma_1(\theta)^3} - \frac{\sigma_{12}(\theta)}{\sigma_1(\theta)m_2(\theta)} + \frac{\sigma_1(\theta)}{2m_2(\theta)^2}m_3(\theta)$$

$$\text{with} \begin{cases} b_1(\theta) := \lim_{n\to+\infty} \mathbb{E}_\theta\big[n\,M_n^{(1)}(\alpha_0)\big] \\ \sigma_{12}(\theta) := \lim_{n\to+\infty} \mathbb{E}_\theta\big[n\,M_n^{(1)}(\alpha_0)\big(M_n^{(2)}(\alpha_0) - m_2(\theta)\big)\big] \\ m_{3,1}(\theta)^3 := \lim_{n\to+\infty} \mathbb{E}_\theta\big[n^2\left(M_n^{(1)}(\alpha_0)\right)^3\big] - 3\sigma_1^2(\theta)b_1(\theta). \end{cases}$$

Proof of Theorem 4. It is sufficient to show that the assumptions of Theorem 4 imply the previous ones of Theorem 3.

- From (B.1), (A.1) holds true with $l := \widetilde{l} + 1$. Indeed let $S_n(\theta) := n(W_n - \widetilde{l}(\theta))$. Thanks to (4) and Markov inequality, one easily obtains $\sup_{\theta\in\Theta}\mathbb{P}_\theta\{W_n \geq l(\theta)\} \leq (1/n)\sup_{\theta\in\Theta}(\mathbb{E}_\theta[S_n(\theta)^2]/n) = O(n^{-1})$.
- (A.2) is directly implied by (B.2), (B.3), (B.4) using Proposition 1, where

  $$B_\theta(u) := \frac{m_{3,1}(\theta)^3}{6\sigma_1(\theta)^3}(1-u^2) - \frac{b_1(\theta)}{\sigma_1(\theta)}.$$

  Similar expression holds for $C_\theta(u)$.

- In a more intricate way, to prove (A.3) under (B.2), (B.3), (B.5), let us define

$$S_n(\theta, p, v) := n\left[M_n^{(1)}(\alpha_0) + \varsigma_p(v,\theta)\big(M_n^{(2)}(\alpha_0) - m_2(\theta)\big) + \frac{\varsigma_p(v,\theta)^2}{2}\big(M_n^{(3)}(\alpha_0) - m_3(\theta)\big)\right]$$

where $\varsigma_p(v,\theta) := v\sigma(\theta)/\sqrt{p}$, so that $S_n(\theta, n, u) = n\widetilde{M_n}(u,\theta)$ (cf. (34)). From (B.5) and using Proposition 1, we can define

$$\sigma_{p,v}(\theta)^2 := \lim_{n\to+\infty} \frac{\mathbb{E}_\theta[S_n(\theta,p,v)^2]}{n},$$

and from (B.3) and using Proposition 1 again, we obtain (A.3) with

$$D_\theta(u) := -\frac{\sigma_{12}(\theta)}{\sigma_1(\theta)^3}\sigma(\theta)u, \quad \text{and} \quad E_\theta(u) := \frac{m_{31}(\theta)^3}{6\sigma_1(\theta)^3}(1-u^2) - \frac{b_1(\theta)}{\sigma_1(\theta)}. \qquad \square$$

6.3 Proof of Theorem 2 of Subsection 5.1

Let us go back to our Markovian model $(\mathcal{M})$ and prove that the assumptions of Theorem 4 hold true whenever the assumptions of Theorem 2 are satisfied.

- Let us define $W_n := (1/n)\sum_{k=1}^n (W(X_{k-1}) + W(X_k))$ and $\widetilde{l}(\theta) := 2\,\mathbb{E}_{\theta,\pi_\theta}[W(X_1)]$, where W is defined in (C.3). Then, using Proposition 3, the Lipschitz condition (B.1) for $j = 3$ is true. Indeed the family $\{\xi_\theta;\ \theta \in \Theta\}$ obviously fulfills the moment domination condition (D_2) with $\xi_\theta(x,y) := W(x) + W(y) - 2\widetilde{l}(\theta)$. In the same way, the remaining part of (B.1) (for $j = 2$) is checked under (27) (which means that the family $\{F^{(3)}(\alpha,\cdot,\cdot);\ \alpha \in \mathcal{A}\}$ fulfills (D_3) and a fortiori (D_2)).
- Firstly, we deduce from Proposition 3 that the part of (B.2) concerning Hypothesis $\mathcal{R}(3)$ is true under (27). Secondly, thanks to Assumption $(\mathcal{S}')$, we deduce from Lemma 3 and Theorem 1 that the part of (B.2) concerning Hypothesis (N-A) is true under Condition (a) of Theorem 2 (see Remark 3 concerning the assumptions of Lemma 3).
- The conditions (B.3) and (B.4) are exactly (C.2).
- We deduce from Proposition 3 that the first point of (B.5) follows from (27). For the second point of (B.5), recall Definition (34) of $\widetilde{M_n}(\theta,u)$, and notice that $\inf_{\theta\in\Theta} m_2(\theta) > 0$ from (C.1) and $\sup_{\theta\in\Theta}\sigma_1(\theta) < +\infty$, which imply that $\sup\{\sigma(\theta)u/\sqrt{n};\ n \geq 1,\ \theta \in \Theta,\ |u| \leq 2\sqrt{\ln n}\} < +\infty$. Thus the family

$$\left\{\sum_{j=1}^{3} \frac{1}{(j-1)!}\left(\frac{\sigma(\theta)}{\sqrt{n}}u\right)^{j-1}\big(F^{(j)}(\alpha_0,\cdot,\cdot) - m_j(\theta)\big);\ n \geq 1, \theta \in \Theta,\ |u| \leq 2\sqrt{\ln n}\right\}$$

obviously fulfills (D_3), and we conclude from Proposition 3 that $\{n\,\widetilde{M_n}(\theta,u);\ n \geq 1,\ \theta \in \Theta,\ |u| \leq 2\sqrt{\ln n}\}$ fulfills Hypothesis $\mathcal{R}(3)$. Finally, thanks to Assumption $(\mathcal{S}')$, we easily check from Lemma 3 and Theorem 1 that the part of (B.5) concerning Hypothesis (N-A) is true under Condition (b) of Theorem 2. $\square$

6.4 Illustration of Theorem 4 in the case of some AR(d) processes

In this subsection, we apply Theorem 4 to linear autoregressive processes of order d, $d \geq 2$. In substance, such a model fulfills all the assumptions of Theorem 2, except the dominated model Condition $(\mathcal{S}')$. Consequently, the non-arithmeticity conditions involved in the assumptions B.2 and B.5 of Theorem 4 cannot be checked using Theorem 1 any more. Here, by reinforcing the assumptions on the density of the noise, we apply the second approach of Subsection 3.3 to study these non-arithmeticity conditions.

Let us consider the following autoregressive process of order $d \geq 1$ on $E := \mathbb{R}^d$:

$$\forall n \geq d, \qquad Y_n := g_1(\theta)Y_{n-1} + \cdots + g_d(\theta)Y_{n-d} + Z_n \tag{37}$$

where the probability distribution of $(Y_0, \ldots, Y_{d-1})$ is denoted by μ_θ and

- $\theta \in \mathbb{R}$ is a parameter,
- $(g_1, \ldots, g_d)$ are given real continuous functions,
- and $(Z_k)_{k \in \mathbb{N}^*}$ are i.i.d. real-valued r.v. independent on $(Y_0, .., Y_{d-1})$ with common distribution which admits some probability density f_Z with respect to μ^{Leb}.

The parameter θ of the observed AR(d) process is assumed to be in a non-empty compact set $\Theta \subset \mathbb{R}$ such that for all $\theta \in \Theta$ the solutions of the equation

$$z^d - g_1(\theta)z^{d-1} - \cdots - g_{d-1}(\theta)z - g_d(\theta) = 0 \tag{38}$$

lie in $D(0,1) := \{z \in \mathbb{C};\ |z| < 1\}$.

Introduce the column vector $X_n := (Y_n, \ldots, Y_{n-d+1})'$ for $n \geq d-1$. Then the process $(X_n)_{n \geq d-1}$ is a Markov chain with the following first-order autoregressive representation

$$\forall n \geq d \qquad X_n = A(\theta)\ X_{n-1} + (Z_n, 0, \ldots, 0)' \tag{39}$$

where

$$A(\theta) := \begin{pmatrix} g_1(\theta) & & \ldots & & g_{d-1}(\theta) & g_d(\theta) \\ 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 & 0 \\ \vdots & & \ddots & & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 & 0 \\ 0 & 0 & \ldots & 0 & 1 & 0 \end{pmatrix}.$$

Assuming that the solutions of the equation (38) lie in $D(0,1)$ is equivalent to assume that the eigenvalues of $A(\theta)$ have moduli strictly less than unity, so that $\|A(\theta)\| < 1$ for all $\theta \in \Theta$ and $\sup_{\theta \in \Theta} \|A(\theta)\| < 1$.

The initial distribution of $(X_n)_{n \geq d-1}$ is μ_θ and its transition kernel Q_θ is given for all Borel set $B \in \mathcal{B}(\mathbb{R}^d)$ by

$$Q_\theta(x, B) = \int_{\mathbb{R}} \mathbf{1}_B \left(A(\theta)x + (z, 0, \cdots, 0)'\right) f_Z(z)\, dz.$$

Note that the transition kernel Q_θ has the following representation:

$$Q_\theta(x, dy) = f_Z\big(y_d - \langle g(\theta), x\rangle\big)\, \mu^{\mathrm{Leb}}(y_d)\, \delta_{x_d}(y_{d-1}) \cdots \delta_{x_2}(y_1), \tag{40}$$

where x denotes the column vector $(x_d, \dots, x_1)'$ and y denotes the column vector $(y_d, \dots, y_1)'$. Then, as already mentioned, the dominated model Condition $(\mathcal{S}')$ is not fulfilled in the multidimensional case ($d \neq 1$).

Next, let us assume that the probability density f_Z of Z_1 fulfills the following conditions:

(A) $\forall z \in \mathbb{R},\ f_Z(z) > 0$;

(B) for all $\theta \in \Theta$, $\mathbb{E}_{\theta,\mu_\theta}[Z_1] = 0$;

(C) for all $\theta \in \Theta$, $\mathbb{E}_{\theta,\mu_\theta}[|Z_1|^{10}] < +\infty$;

(D) f_Z is 4-time-differentiable on $\mathbb{R}$;

(E) for $j = 1, \dots, 4$, $f_Z^{(j)}/f_Z$ is a bounded function;

(F) for all $9 < \gamma \leq 10$, there exists $0 < \beta \leq 1 - 1/\gamma$ such that, for all $x_0 \in \mathbb{R}$, there exists a neighborhood V_{x_0} of x_0 and a positive measurable function q_{x_0} satisfying

$$\int_{\mathbb{R}} \big(1 + |y|\big)^\beta q_{x_0}(y)\, dy < \infty \quad \text{and} \quad \forall y \in \mathbb{R},\ \forall t \in V_{x_0},\ f_Z(y+t) \leq q_{x_0}(y).$$

Actually, under (E), it is sufficient to assume in (C) that there exists some $\varepsilon > 0$ such that for all $\theta \in \Theta$, $\mathbb{E}_{\theta,\mu_\theta}[Z_1^{9+\varepsilon}] < +\infty$. Furthermore, note that Assumption (E) can be relaxed provided that the order of the moment of Z_1 is increased. However, Assumption (E) as above is satisfied in several interesting models and thus it does not need relaxing. Notice that Götze and Hipp [GH94, th 1.5] assume that, under (E), Z_1 admits a moment of order 15.

Finally, the vector $g := (g_1, \dots, g_d)'$ is supposed to have the following properties:

(G) $\theta \mapsto\ (g_1(\theta), \dots, g_d(\theta)) \in \mathbb{R}^d$ is 4-time-continuously-differentiable on Θ;

(H) $\inf_{\theta\in\Theta} g_1^{(1)}(\theta) > 0$.

Let us define for all $9 < \gamma \leq 10$

$$\forall x \in \mathbb{R}^d, \quad V(x) := 1 + \|x\|^\gamma. \tag{41}$$

Lemma 8. *Under the previous conditions, $(X_n)_{n\geq d-1}$ belongs the class of Models $(\mathcal{M})$ with the function V defined in (41).*

Proof of Lemma 8. Under (C), one has for all $\theta \in \Theta$ and $x \in \mathbb{R}^d$,

$$\frac{Q_\theta V(x)}{V(x)} = \int_{\mathbb{R}} \frac{V\big(A(\theta)x + (z, 0, \dots, 0)'\big)}{V(x)}\, f_Z(z)\, dz \leq \int_{\mathbb{R}} \frac{1 + \big(\|A(\theta)\|\ \|x\| + |z|\big)^\gamma}{V(x)}\, f_Z(z) dz.$$

By Fatou lemma,

$$\limsup_{\|x\|\to\infty}\left(\sup_{\theta\in\Theta}\frac{Q_\theta V(x)}{V(x)}\right)\le\sup_{\theta\in\Theta}\|A(\theta)\|^\gamma<1.$$

Next, pick $\varrho\in(\sup_{\theta\in\Theta}\|A(\theta)\|^\gamma\,,\,1)$. There exists $s>0$ such that for all $\|x\|>s$ and $\theta\in\Theta$, $Q_\theta V(x)\le\varrho V(x)$. Set $S:=\{x\in\mathbb{R}^d;\ \|x\|\le s\}$. Note that

$$\forall\theta\in\Theta,\ \forall x\in S,\quad Q_\theta V(x)\ \le\ \varsigma:=\sup_{\theta\in\Theta}\int_{\mathbb{R}}\big(1+\big(\|A(\theta)\|\ \|s\|+|z|\big)^\gamma\big)f_Z(z)dz<+\infty,$$

so that

$$\forall\theta\in\Theta,\ \forall x\in\mathbb{R}^d,\quad Q_\theta V(x)\ \le\ \varrho\ V(x)+\varsigma.$$

Finally, since Condition (A) holds true, it is easily checked that $(X_k)_{k\ge0}$ is μ_d^{Leb}-irreducible, aperiodic and fulfills the drift-criterion [MT93] uniformly in $\theta\in\Theta$. □

Set $e_1:=(1,0,\dots,0)'\in\mathbb{R}^d$. Then, let us consider the MLE $(\hat\theta_n)_{n\in\mathbb{N}^*}$ of the parameter θ $(\alpha_0(\theta)\equiv\theta)$. We have

$$\forall n\ge d,\quad\langle e_1,X_n\rangle=\langle g(\theta),X_{n-1}\rangle+Z_n.$$

Maximum likelihood estimation requires to deal with the following function F

$$\forall(x,y)\in\mathbb{R}^d\times\mathbb{R}^d,\ \forall\alpha\in\Theta,\quad F(\alpha,x,y):=-\ln f_Z\big(\langle e_1,y\rangle-\langle g(\alpha),x\rangle\big),$$

and the empirical mean functional

$$\forall n\in\mathbb{N}^*,\ \forall\alpha\in\Theta,\quad M_n(\alpha):=-\frac{1}{n}\sum_{k=1}^n\ln f_Z\big(\langle e_1,X_k\rangle-\langle g(\alpha),X_{k-1}\rangle\big).$$

Proposition 4. *Assume that the previous assumptions on the model hold true and that the MLE $(\hat\theta_n)_{n\in\mathbb{N}^*}$ of the parameter θ associated with $(X_n)_{n\ge d-1}$ satisfies the uniform consistency Property (HYP.2). In addition, assume that the initial probability measure satisfies $\sup_{\theta\in\Theta}\mu_\theta(V)<+\infty$ (that is Property (12)) and that the following conditions hold true:*

1. *$\forall\alpha\in\mathcal{A}$, $\forall j=1,2$, $F^{(j)}(\alpha,x,y)$ fulfills (N-L)' as defined page 12*
2. *$\forall\alpha\in\mathcal{A}$, $\forall v\in\mathbb{R}$, $F^{(1)}(\alpha,x,y)+vF^{(2)}(\alpha,x,y)+(v^2/2)F^{(3)}(\alpha,x,y)$ fulfills (N-L)'.*

Then all the conclusions of Theorem 2 are true.

Proof of Proposition 4. It is easily checked that the family $\{F^{(j)}(\alpha,x,y);\ \alpha\in\mathcal{A},\ j=1,2,3\}$ fulfills the moment domination condition (D_3) (i.e. (27)) mainly thanks to (E) and $\gamma>3\times3$. Next, we claim that Conditions (C.1) to (C.5) of Theorem 2 are satisfied. Indeed, concerning Conditions (C.1) and (C.2), and recalling that $(\pi_\theta)_{\theta\in\Theta}$ denotes the invariant probability of the Markov chain $(X_n)_{n\ge d-1}$, we have

- $m_1(\theta)=\mathbb{E}_{\theta,\pi_\theta}[\langle g^{(1)}(\theta),X_0\rangle]\mathbb{E}_{\theta,\pi_\theta}[f_Z^{(1)}(Z_1)/f_Z(Z_1)]\equiv0$;

  $m_2(\theta)=\mathbb{E}_{\theta,\pi_\theta}[\langle g^{(1)}(\theta),X_0\rangle^2]\ \mathbb{E}_{\theta,\pi_\theta}\left[f_Z^{(1)}(Z_1)^2/f_Z(Z_1)^2\right]\ge g_1^{(1)}(\theta)^2\mathbb{E}_{\theta,\pi_\theta}[Z_1^2]$, which implies that $\inf_{\theta\in\Theta}m_2(\theta)>0$.

- $\sigma_1(\theta)^2 = m_2(\theta)$, hence one has $\inf_{\theta\in\Theta} \sigma_1(\theta) > 0$;
  $\sigma_2(\theta)^2 \geq C$ where $C > 0$ depends on the variances of Z_1 and Z_1^2.

Conditions (C.4) and (C.5) are obviously satisfied. Concerning (C.3), use the fact that the family $\{F^{(4)}(\alpha, x, y);\ \alpha \in \mathcal{A}\}$ fulfills (D_2) (this statement holds true mainly thanks to Assumption (E) and $\gamma > 2 \times 4$).

By using the previous facts and proceeding as in the proof of Theorem 2 (see Subsection 6.3), one can see that all the assumptions of Theorem 4 but those concerning Hypothesis (N-A) are fulfilled. Consequently, to deduce Proposition 4 from Theorem 4, it only remains to establish that the characteristic functions of the following families $(\xi_p)_{p\in\mathcal{P}}$ (involved in Assumptions B.2 and B.5 of Theorem 4) satisfy Hypothesis (N-A):

(a) $\{F^{(j)}(\alpha, x, y);\ \alpha \in \mathcal{A}\}$ with $j = 1, 2$,

(b) $\{F^{(1)}(\alpha, x, y) + \upsilon F^{(2)}(\alpha, x, y) + (\upsilon^2/2)F^{(3)}(\alpha, x, y);\ \alpha \in \mathcal{A},\ \upsilon \in I\!R\}$.

To that effect, we make use of the second approach of Subsection 3.3. Below, (i), (ii) and (iii) refer to the conditions introduced in Subsection 3.1:

Fact1. Families (a)-(b) satisfy Condition (N-L). Indeed, thanks to Conditions $(\mathcal{C})$ and $(\mathcal{S})$ (use the fact that $f_Z > 0$ to check that $(\mathcal{S})$ holds true), we deduce from Lemma 3 that Fact1 follows from Assumptions 1. and 2. of this Proposition 4.

Fact2. The Fourier operators of Families (a)-(b) satisfy Condition (i). Indeed Assumptions 1. and 2. of Lemma 1 are fulfilled (see the comments after Lemma 1 concerning Assumption 1. and Property (17) concerning Assumption 2.). Then, using Lemma 1, Fact2 follows from Fact1.

Fact3. The Fourier operators of Families (a)-(b) satisfy (ii)-(iii). Indeed notice that the family $\{F^{(j)}(\alpha, x, y);\ \alpha \in \mathcal{A},\ j = 1, \ldots, 4\}$ satisfies (D_0), and consequently, the assumptions of Proposition 2 are fulfilled (see Lemma 9 below and apply it to the case where $(\xi_p)_{p\in\mathcal{P}}$ is any of the above Families (a)-(b)). Then, using Proposition 2, Fact3 follows from Fact2.

Fact4. Finally the Fourier operators of Families (a)-(b) satisfy Hypothesis (N-A'), and so (N-A) (see Lemma 2 and see also the end of Section 2).

The proof of Proposition 4 is now complete, provided that we give the proof of the next lemma. □

Lemma 9. *Assume that $(\xi_p)_{p\in\mathcal{P}}$ and its derivative with respect to p fulfill (D_{m_0}) with some $m_0 \in I\!N$. Let $Q_{\theta,p}(t)$ be the Fourier operator defined by (8) where the transition kernel Q_θ is the particular one given in (40). Then, Conditions (C1) and (C2) of Proposition 2 hold true.*

Proof of Lemma 9. Let us prove that the family $(\xi_p)_{p\in\mathcal{P}}$ verifies the assumptions of Proposition 2 with $\mathcal{B} := \mathcal{B}_\beta \hookrightarrow \widetilde{\mathcal{B}} := \mathcal{B}_1$ where β is defined in Assumption (F) (see page 15 for the definition of the spaces).

First, Condition (C1) of Proposition 2 is exactly (D-F). Then, concerning Condition (C2) of Proposition 2, let us prove that the following properties are valid:

1. the map $t \mapsto Q_{\theta,p}(t)$ is continuous from $I\!R$ into $\mathcal{L}(\mathcal{B}_\beta, \mathcal{B}_1)$ uniformly in $(\theta, p) \in \Theta \times \mathcal{P}$;
2. for all $t \in I\!R$, the map $(\theta, p) \mapsto Q_{\theta,p}(t)$ is continuous from $\Theta \times \mathcal{P}$ into $\mathcal{L}(\mathcal{B}_\beta, \mathcal{B}_1)$.

Then one obviously has $\|Q_{\theta,p}(t) - Q_{\theta_0,p_0}(t_0)\|_{\mathcal{B}_\beta,\mathcal{B}_1} \to 0$ when $(t, \theta, p) \to (t_0, \theta_0, p_0)$, which completes the proof of Lemma 9.

Let us first introduce the real number $E_\beta := c_\beta \kappa_\beta + b_1$, where $b_1 := \sup_{\theta \in \Theta} \pi_\theta(V) < +\infty$ from (VG1) and κ_β and c_β are defined in (D-F). Then, using (D-F), $V \geq 1$ and $\beta > 0$, we obtain

$$\forall \theta \in \Theta, \qquad Q_\theta V^\beta \leq E_\beta V^\beta.$$

On the other hand, let us state the following obvious inequality:

$$\forall a \in I\!R, \quad |e^{ia} - 1| \leq \min(2, |a|) \leq 2|a|^\alpha.$$

Now recall that $0 < \beta < 1$ and let us define $0 < \alpha \leq 1$ such that $\beta + \alpha/(m_0 + \varepsilon) \leq 1$ where $\varepsilon > 0$ is defined in (D_{m_0}).

1) Let us define $\Delta := Q_{\theta,p}(t) - Q_{\theta,p}(t_0)$. One has for all $f \in \mathcal{B}_\beta$ and $x \in E$:

$$\begin{aligned}
|\Delta f(x)| &\leq \int_E \left| e^{it\xi_p(x,y)} - e^{it_0\xi_p(x,y)} \right| |f(y)|\, Q_\theta(x, dy) \\
&\leq 2\,|t - t_0|^\alpha\, \|f\|_{\mathcal{B}_\beta} \int_E |\xi_p(x, y)|^\alpha\, V(y)^\beta Q_\theta(x, dy) \\
&\leq 2\, C_\xi^{\frac{\alpha}{m_0+\varepsilon}}\, |t - t_0|^\alpha\, \|f\|_{\mathcal{B}_\beta} \int_E (V(x) + V(y))^{\frac{\alpha}{m_0+\varepsilon}}\; V(y)^\beta Q_\theta(x, dy) \\
&\leq 2^{1+\frac{\alpha}{m_0+\varepsilon}}\, C_\xi^{\frac{\alpha}{m_0+\varepsilon}}\, |t - t_0|^\alpha\, \|f\|_{\mathcal{B}_\beta} \left(V(x)^{\frac{\alpha}{m_0+\varepsilon}}\, Q_\theta V^\beta(x) + Q_\theta V^{\beta+\frac{\alpha}{m_0+\varepsilon}}(x) \right) \\
&\leq 2^{1+\frac{\alpha}{m_0+\varepsilon}}\, C_\xi^{\frac{\alpha}{m_0+\varepsilon}}\, |t - t_0|^\alpha\, \|f\|_{\mathcal{B}_\beta} \left(E_\beta + E_{\beta+\frac{\alpha}{m_0+\varepsilon}} \right) V(x)^{\beta+\frac{\alpha}{m_0+\varepsilon}}
\end{aligned}$$

from which we deduce $\|\Delta f\|_{\mathcal{B}_1} \leq D_\xi\, |t - t_0|^\alpha\, \|f\|_{\mathcal{B}_\beta}$ where D_ξ does not depend on (θ, p).

2) In the same way, let us define $\Delta := Q_{\theta,p}(t) - Q_{\theta_0,p_0}(t)$. Let x denote some d-dimensional column vector $(x_d, \ldots, x_1)'$ and $\underline{x}_{d-1}$ denote the associated $(d-1)$-dimensional column vector

$(x_d,\ldots,x_2)'$. We have for all $f\in\mathcal{B}_\beta$ and $x\in E$:

$$\begin{aligned}
\Delta f(x) &= \int_{I\!R} \exp\left(it\xi_p\left(x,\begin{pmatrix}\langle x,g(\theta)\rangle+z\\ \underline{x}_{d-1}\end{pmatrix}\right)\right) f\begin{pmatrix}\langle x,g(\theta)\rangle+z\\ \underline{x}_{d-1}\end{pmatrix} f_Z(z)\,dz\\
&\quad -\int_{I\!R} \exp\left(it\xi_{p_0}\left(x,\begin{pmatrix}\langle x,g(\theta_0)\rangle+z\\ \underline{x}_{d-1}\end{pmatrix}\right)\right) f\begin{pmatrix}\langle x,g(\theta_0)\rangle+z\\ \underline{x}_{d-1}\end{pmatrix} f_Z(z)\,dz\\
&= \int_{I\!R} \exp\left(it\xi_p\left(x,\begin{pmatrix}y\\ \underline{x}_{d-1}\end{pmatrix}\right)\right) f\begin{pmatrix}y\\ \underline{x}_{d-1}\end{pmatrix} f_Z(y-\langle x,g(\theta)\rangle)\,dy\\
&\quad -\int_{I\!R} \exp\left(it\xi_{p_0}\left(x,\begin{pmatrix}y\\ \underline{x}_{d-1}\end{pmatrix}\right)\right) f\begin{pmatrix}y\\ \underline{x}_{d-1}\end{pmatrix} f_Z(y-\langle x,g(\theta_0)\rangle)\,dy\\
|\Delta f(x)| &\leq \left|\int_{I\!R} f\begin{pmatrix}y\\ \underline{x}_{d-1}\end{pmatrix}\left(f_Z(y-\langle x,g(\theta)\rangle)-f_Z(y-\langle x,g(\theta_0)\rangle)\right)dy\right|\\
&\quad +2|t|^\alpha\int_{I\!R}\left|(\xi_p-\xi_{p_0})\left(x,\begin{pmatrix}y\\ \underline{x}_{d-1}\end{pmatrix}\right)\right|^\alpha \left|f\begin{pmatrix}y\\ \underline{x}_{d-1}\end{pmatrix}\right| f_Z(y-\langle x,g(\theta_0)\rangle)\,dy.
\end{aligned}$$

Thus one has $|\Delta f(x)|\leq \|f\|_{\mathcal{B}_\beta}\,(|\theta-\theta_0|\,I_1+|t|^\alpha\,|p-p_0|^\alpha\,I_2)$ where, thanks to differentiation under the integral sign and Assumptions (E)-(F), I_1 satisfies for some $\widetilde{\theta}\in I\!R$ such that $|\widetilde{\theta}-\theta|\leq|\widetilde{\theta}-\theta_0|$

$$\begin{aligned}
I_1\leq\ & \sup_{\widetilde{\theta}\in\Theta}|\langle x,g^{(1)}(\widetilde{\theta})\rangle|\,\sup_{z\in I\!R}\left|\frac{f_Z^{(1)}(z)}{f_Z(z)}\right|\int_{I\!R}V\begin{pmatrix}y\\ \underline{x}_{d-1}\end{pmatrix}^\beta f_Z(y-\langle x,g(\widetilde{\theta})\rangle)\,dy\\
&=\sup_{\widetilde{\theta}\in\Theta}|\langle x,g^{(1)}(\widetilde{\theta})\rangle|\,\sup_{z\in I\!R}\left|\frac{f_Z^{(1)}(z)}{f_Z(z)}\right| Q_{\widetilde{\theta}}V^\beta(x)\\
&\leq E_\beta\,V(x)^{\frac{1}{\gamma}+\beta}\,\sup_{\widetilde{\theta}\in\Theta}\|g^{(1)}(\widetilde{\theta})\|\,\sup_{z\in I\!R}\left|\frac{f_Z^{(1)}(z)}{f_Z(z)}\right|
\end{aligned}$$

and where I_2 satisfies on the other hand

$$\begin{aligned}
I_2\leq\ & 2\int_{I\!R}\left(V(x)+V\begin{pmatrix}y\\ \underline{x}_{d-1}\end{pmatrix}\right)^{\frac{\alpha}{m_0+\varepsilon}} V\begin{pmatrix}y\\ \underline{x}_{d-1}\end{pmatrix}^\beta f_Z(y-\langle x,g(\theta_0)\rangle)\,dy\\
&=\ 2^{1+\frac{\alpha}{m_0+\varepsilon}}\left(V(x)^{\frac{\alpha}{m_0+\varepsilon}}Q_{\theta_0}V^\beta(x)+Q_{\theta_0}V^{\beta+\frac{\alpha}{m_0+\varepsilon}}(x)\right)\\
&\leq\ 2^{1+\frac{\alpha}{m_0+\varepsilon}}\left(E_\beta+E_{\beta+\frac{\alpha}{m_0+\varepsilon}}\right)V^{\beta+\frac{\alpha}{m_0+\varepsilon}}(x).
\end{aligned}$$

Since $0<\beta\leq 1-1/\gamma$, one has $\|\Delta f\|_{\mathcal{B}_1}\leq D'_\xi\,(|\theta-\theta_0|+|t|^\alpha\,|p-p_0|^\alpha)\,\|f\|_{\mathcal{B}_\beta}$. □

A Proof of Theorem 3 of Subsection 6.1

The investigation of the case $|u|>2\sqrt{\ln n}$ is similar to the one of [HLP], so that the details are omitted. By contrast, the case $|u|\leq 2\sqrt{\ln n}$ is quite different. First let us introduce for all $\theta\in\Theta$ and $u\in I\!R$, $|u|\leq 2\sqrt{\ln n}$

$$\tau=\tau_n(u,\theta):=\alpha_0+\frac{\sigma(\theta)}{\sqrt{n}}u\quad\text{and}\quad \varsigma_n(u,\theta):=\frac{\sigma(\theta)}{\sqrt{n}}u=\tau-\alpha_0.$$

For the sake of simplicity, let us define for all $\theta \in \Theta$ and $u \in I\!R$, $|u| \leq 2\sqrt{\ln n}$

$$P_{n,\theta}(u) := \mathbb{P}_\theta \left\{ \frac{\sqrt{n}}{\sigma(\theta)} (\hat{\alpha}_n - \alpha_0) \leq u \right\} = \mathbb{P}_\theta \left\{ \hat{\alpha}_n \leq \tau \right\}, \quad Q_{n,\theta}(u) := \mathbb{P}_\theta \left\{ M_n^{(1)}(\tau) \geq 0 \right\}.$$

At a first stage we prove that

$$\sup_{\theta\in\Theta} \sup_{|u|\leq 2\sqrt{\ln n}} |P_{n,\theta}(u) - Q_{n,\theta}(u)| = o(n^{-\frac{1}{2}}) \tag{A}$$

and then we determine A_θ such that

$$\sup_{\theta\in\Theta} \sup_{|u|\leq 2\sqrt{\ln n}} \left| Q_{n,\theta}(u) - \Big(\mathcal{N}(u) + \eta(u) n^{-\frac{1}{2}} A_\theta(u) \Big) \right| = o(n^{-\frac{1}{2}}) \tag{B}$$

to complete the proof of Theorem 3.

Let us prove that (A) holds true. It follows from (HYP.1) that there exists some real r.v. $\widetilde{\alpha}'_n$ such that $|\widetilde{\alpha}'_n - \tau| < |\hat{\alpha}_n - \tau|$ and $0 = M_n^{(1)}(\tau) + (\hat{\alpha}_n - \tau)\, M_n^{(2)}(\widetilde{\alpha}'_n)$. Next, introducing the event $\{M_n^{(2)}(\widetilde{\alpha}'_n) > 0\}$ and its complement, one has

$$P_{n,\theta}(u) = \mathbb{P}_\theta\big\{M_n^{(1)}(\tau) \geq 0, M_n^{(2)}(\widetilde{\alpha}'_n) > 0\big\} + \mathbb{P}_\theta\big\{\hat{\alpha}_n \leq \tau, M_n^{(2)}(\widetilde{\alpha}'_n) \leq 0\big\},$$

so that

$$|P_{n,\theta}(u) - Q_{n,\theta}(u)| \leq 2\ \mathbb{P}_\theta\big\{M_n^{(2)}(\widetilde{\alpha}'_n) \leq 0\big\}.$$

Introducing the events $\{M_n^{(2)}(\widetilde{\alpha}'_n) < M_n^{(2)}(\alpha_0) - |\widetilde{\alpha}'_n - \alpha_0| l(\theta)\}$, $\{M_n^{(2)}(\alpha_0) \leq m_2(\theta)/2\}$ and their complements, where the function $l(\cdot)$ is defined in (A.1), one has

$$\mathbb{P}_\theta\big\{M_n^{(2)}(\widetilde{\alpha}'_n) \leq 0\big\} \leq P_1 + P_2 + P_3,$$

where $(P_i)_{i=1,2,3}$ denote

$$\begin{aligned}
P_1 &:= \sup_{\theta\in\Theta} \sup_{|u|\leq 2\sqrt{\ln n}} \mathbb{P}_\theta \left\{ M_n^{(2)}(\widetilde{\alpha}'_n) < M_n^{(2)}(\alpha_0) - |\widetilde{\alpha}'_n - \alpha_0| l(\theta) \right\} \\
P_2 &:= \sup_{\theta\in\Theta} \mathbb{P}_\theta \left\{ M_n^{(2)}(\alpha_0) \leq \frac{m_2(\theta)}{2} \right\} \\
P_3 &:= \sup_{\theta\in\Theta} \sup_{|u|\leq 2\sqrt{\ln n}} \left\{ \frac{m_2(\theta)}{2} - |\widetilde{\alpha}'_n - \alpha_0| l(\theta) < M_n^{(2)}(\widetilde{\alpha}'_n) \leq 0 \right\}.
\end{aligned}$$

- Introducing the event $\{W_n \geq l(\theta)\}$ and its complement, it is easy to check from (A.1) that $P_1 = o(n^{-1/2})$.

- One has $P_2 \leq \mathbb{P}_\theta \left\{ (\sqrt{n}/\sigma_2(\theta))\big(M_n^{(2)}(\alpha_0) - m_2(\theta)\big) \leq -b\sqrt{n} \right\}$ where $b := \inf_{\theta\in\Theta} m_2(\theta)/2\sigma_2(\theta)$, $b > 0$ from (A.2), which implies $P_2 = o(n^{-1/2})$.

- Introducing the event $\{|\widetilde{\alpha}'_n - \alpha_0| \geq m_2(\theta)/2l(\theta)\}$ and its complement, it is easily checked that
$$P_3 \leq \sup_{\theta\in\Theta} \sup_{|u|\leq 2\sqrt{\ln n}} \mathbb{P}_\theta \left\{|\widetilde{\alpha}'_n - \alpha_0| \geq m_2(\theta)/2l(\theta)\right\}.$$
Furthermore $\widetilde{\alpha}'_n$ satisfies $|\widetilde{\alpha}'_n - \alpha_0| \leq |\hat{\alpha}_n - \alpha_0| + 2|\tau - \alpha_0|$, where $\sup_{\theta\in\Theta}\sup_{|u|\leq 2\sqrt{\ln n}} |\tau - \alpha_0| \to 0$ when $n \to +\infty$ (recall that $\sup_{\theta\in\Theta} \sigma(\theta) < +\infty$). Thus $P_3 \leq \sup_{\theta\in\Theta} \mathbb{P}_\theta\{|\hat{\alpha}_n - \alpha_0| \geq d\}$ for n sufficiently large, and where the real number d is defined by $d := \inf_{\theta\in\Theta} m_2(\theta)/(4l(\theta)) > 0$, so that $P_3 = o(n^{-1/2})$.

Therefore the estimate (A) holds true.

In a second and last step, let us determine A_θ such that (B) holds true. There exists some real r.v. $\widetilde{\alpha}''_n$ such that $|\widetilde{\alpha}''_n - \alpha_0| < |\tau - \alpha_0|$ and

$$M_n^{(1)}(\tau) = M_n^{(1)}(\alpha_0) + \varsigma_n(u,\theta)\, M_n^{(2)}(\alpha_0) + \frac{\varsigma_n(u,\theta)^2}{2}\, M_n^{(3)}(\widetilde{\alpha}''_n).$$

Let us introduce the r.v.

$$Z_n(u,\theta) := M_n^{(1)}(\alpha_0) + \varsigma_n(u,\theta)\, M_n^{(2)}(\alpha_0) + \frac{\varsigma_n(u,\theta)^2}{2}\, M_n^{(3)}(\alpha_0),$$

the event $C_{n,\theta} := \{W_n < l(\theta)\}$ and the positive number $c = c_n(u,\theta) := |\varsigma_n(u,\theta)|^3\, l(\theta)/2$, where the r.v. W_n and the function $l(\cdot)$ are defined in (A.1).

Consider the following events

$$\begin{aligned} B^{1-}_{n,u,\theta} &:= \{Z_n(u,\theta) - c \geq 0\}, & B^{2-}_{n,u,\theta} &:= B^{1-}_{n,u,\theta} \cap C_{n,\theta} \\ \widetilde{B^{1}_{n,u,\theta}} &:= \left\{M_n^{(1)}(\tau) \geq 0\right\}, & \widetilde{B^{2}_{n,u,\theta}} &:= \widetilde{B^{1}_{n,u,\theta}} \cap C_{n,\theta} \\ B^{1+}_{n,u,\theta} &:= \{Z_n(u,\theta) + c \geq 0\}, & B^{2+}_{n,u,\theta} &:= B^{1+}_{n,u,\theta} \cap C_{n,\theta} \end{aligned}$$

and notice that $Q_{n,\theta}(u) = \mathbb{P}_\theta\{\widetilde{B^{1}_{n,u,\theta}}\}$ and the following facts

- since $\sup_{\theta\in\Theta} \mathbb{P}_\theta\{C^c_{n,\theta}\} = o(n^{-1/2})$ from (A.1), one has $\sup_{\theta\in\Theta}\sup_{|u|\leq 2\sqrt{\ln n}} |Q_{n,\theta}(u) - \mathbb{P}_\theta\{\widetilde{B^{2}_{n,u,\theta}}\}| = o(n^{-1/2})$;

- one obviously has $M_n^{(1)}(\tau) = Z_n(u,\theta) + (\varsigma_n(u,\theta)^2/2)\,(M_n^{(3)}(\widetilde{\alpha}''_n) - M_n^{(3)}(\alpha_0))$, and hence from (A.1), one has
$$\forall\theta\in\Theta,\ \forall u \in I\!R, \qquad B^{2-}_{n,u,\theta} \subset \widetilde{B^{2}_{n,u,\theta}} \subset B^{2+}_{n,u,\theta}$$

- again since $\sup_{\theta\in\Theta} \mathbb{P}_\theta\{C^c_{n,\theta}\} = o(n^{-1/2})$, one obtains $\sup_{\theta\in\Theta}\sup_{|u|\leq 2\sqrt{\ln n}} |\mathbb{P}_\theta\{B^{2+}_{n,u,\theta}\} - \mathbb{P}_\theta\{B^{1+}_{n,u,\theta}\}| = o(n^{-1/2})$ and $\sup_{\theta\in\Theta}\sup_{|u|\leq 2\sqrt{\ln n}} |\mathbb{P}_\theta\{B^{2-}_{n,u,\theta}\} - \mathbb{P}_\theta\{B^{1-}_{n,u,\theta}\}| = o(n^{-1/2})$.

Then it only remains to determine A_θ such that

$$\sup_{\theta\in\Theta}\sup_{|u|\le 2\sqrt{\ln n}}|\mathbb{P}_\theta\{B^{1\pm}_{n,u,\theta}\} - \Big(\mathcal{N}(u)+\eta(u)n^{-\frac{1}{2}}A_\theta(u)\Big)| = o(n^{-\frac{1}{2}}).$$

Let us introduce

$$\begin{aligned}\Delta_n^{\pm}(u,\theta) &:= \frac{\sqrt{n}}{\sigma_{n,u}(\theta)}\left[m_2(\theta)\varsigma_n(u,\theta)+\frac{\varsigma_n(u,\theta)^2}{2}m_3(\theta)\pm c\right]-u\\ &= u\left(\sigma_{n,u}(\theta)^{-1}\left[\sigma_1(\theta)+\sigma(\theta)\frac{m_3(\theta)}{2}\varsigma_n(u,\theta)\pm\sigma(\theta)\,l(\theta)\frac{\varsigma_n(u,\theta)^2}{2}\right]-1\right)\end{aligned}$$

so that

$$\mathbb{P}_\theta\{B^{1\pm}_{n,u,\theta}\} = 1-\mathbb{P}_\theta\left\{\frac{\sqrt{n}}{\sigma_{n,u}(\theta)}\widetilde{M_n}(u,\theta) < -u-\Delta_n^{\pm}(u,\theta)\right\}.$$

From the last property of (A.3) applied to $v=-u-\Delta_n^{\pm}(u,\theta)$, we obtain

$$\begin{aligned}\sup_{\theta\in\Theta}\sup_{|u|\le 2\sqrt{\ln n}}\Big|&\mathbb{P}_\theta\{B^{1\pm}_{n,u,\theta}\}\le -u-\Delta_n^{\pm}(u,\theta)\Big\}\\ &-\mathcal{N}(u+\Delta_n^{\pm}(u,\theta))+\eta(u+\Delta_n^{\pm}(u,\theta))n^{-\frac{1}{2}}E_\theta(-u-\Delta_n^{\pm}(u,\theta))\Big| = o(n^{-\frac{1}{2}}).\end{aligned}$$

From the first property of (A.3), both $\Delta_n^{+}(u,\theta)$ and $\Delta_n^{-}(u,\theta)$ admit the following expansion:

$$\sup_{\theta\in\Theta}\sup_{|u|\le 2\sqrt{\ln n}}\left|\Delta_n^{\pm}(u,\theta)-\left(\sigma_1(\theta)D_\theta(u)+\frac{\sigma(\theta)^2}{2\sigma_1(\theta)}m_3(\theta)\ u\right)u\ n^{-\frac{1}{2}}\right| = o(n^{-\frac{1}{2}}),$$

and hence

$$\sup_{\theta\in\Theta}\sup_{|u|\le 2\sqrt{\ln n}}|\mathbb{P}_\theta\{B^{1+}_{n,u,\theta}\}-\mathbb{P}_\theta\{B^{1-}_{n,u,\theta}\}| = o(n^{-\frac{1}{2}}).$$

Finally we define the polynomial function A_θ as follows to obtain (36):

$$\begin{aligned}\sup_{\theta\in\Theta}\sup_{|u|\le 2\sqrt{\ln n}}\Big|&\big(\mathcal{N}(u)+\eta(u)n^{-\frac{1}{2}}A_\theta(u)\big)-\\ &\big(\mathcal{N}(u+\Delta_n^{+}(u,\theta))-\eta(u+\Delta_n^{+}(u,\theta))n^{-\frac{1}{2}}E_\theta(-u-\Delta_n^{+}(u,\theta))\big)\Big| = o(n^{-\frac{1}{2}}).\end{aligned}$$ □

B The regeneration method versus Fourier techniques (combined with operator perturbation theorems)

This appendix is to convince that the regeneration method cannot easily be applied to our context.

Recall that in this paper we have studied the behavior of $\mathbb{P}_{\theta,\mu_\theta}\left\{S_n(p)/(\sigma_{\theta,p}\sqrt{n})\le u\right\}$ uniformly in $(\theta,p)\in\Theta\times\mathcal{P}$ and $u\in I\!R$ thanks to Fourier techniques, where $S_n(p)$ is defined by

(1), i.e. $S_n(p) := \sum_{k=1}^n \xi_p(X_{k-1}, X_k)$. In this appendix, we want to highlight the drawbacks of using the regeneration method to investigate the same issue.

Let us mention that there are mainly two constraints we must deal with: limit theorems must be obtained uniformly in both the model parameter θ and the technical parameter p, and they concern bivariate functions ξ_p (with a view to making statistical inference).

To the best of our knowledge, on the one hand, limit theorems with an effective control of the constants which are obtained thanks to the regeneration method only deal with Berry-Esseen theorem and only concern univariate additive functionals $\sum_{k=1}^n \xi_p(X_k)$ (cf. [BC11, DM06]). On the other hand, limit theorems concerning bivariate additive functionals which are obtained thanks to the regeneration method do not control the constants (cf. [Jen89]).

There is one way to bypass the bivariate constraint: we may consider directly the double sequence $(Y_n)_{n\geq 1}$ where $Y_n := (X_{n-1}, X_n)$. However, as explained in §B.1, this choice induces too strong restrictions on the model. In the same way, considering the simple sequence $(X_n)_{n\geq 0}$ as in §B.2 induces very restrictive conditions on the initial probability, excepted for models with an atom. Moreover note that, in atomic models, the uniform control for bivariate functionals has not been investigated by regenerative methods.

In the sequel, we drop the parameters (θ, p) for the sake of notational simplicity. Recall that $(X_n)_{n\geq 0}$ is a Markov chain with state space $(E, \mathcal{E})$ and transition kernel Q.

B.1 Application of the results of [BC11] to the sequence $(Y_n)_{n\geq 1}$

Before explaining why applying the usual regeneration method to the sequence $(Y_n)_{n\geq 1} = (X_{n-1}, X_n)_{n\geq 1}$ induces restrictions, let us briefly explain how it works. The sequence $(Y_n)_{n\geq 1}$ is a Markov chain with state space $(E\times E, \mathcal{E}\otimes\mathcal{E})$ and transition kernel P defined by:

$$\forall F \in \mathcal{E}\otimes\mathcal{E}, \quad P\big((x,y);F\big) := \int_E \mathbf{1}_F(y,z)Q(y,dz).$$

Let us recall that $(Y_n)_{n\geq 1}$ is said to be ψ-irreducible for some positive measure ψ on $(E\times E, \mathcal{E}\otimes\mathcal{E})$ if, for all $F\in\mathcal{E}\otimes\mathcal{E}$, we have:

$$\psi(F) > 0 \;\Rightarrow\; \forall (x,y)\in E\times E, \quad \sum_{n=2}^{+\infty} \mathbb{P}_{(x,y)}\big\{Y_n \in F\big\} > 0,$$

and $(Y_n)_{n\geq 1}$ is said to be Harris-recurrent under the ψ-irreducibility assumption if, for all $F\in\mathcal{E}\otimes\mathcal{E}$:

$$\psi(F) > 0 \;\Rightarrow\; \forall (x,y)\in E\times E, \quad \mathbb{P}_{(x,y)}\left\{\sum_{n=1}^{+\infty} \mathbf{1}_F(Y_n) = +\infty\right\} = 1.$$

B.1.1 Discussion under existence of an atom for $(Y_n)_{n\geq 1}$

Under the ψ-irreducibility assumption, we suppose that $(Y_n)_{n\geq 1}$ is Harris-recurrent and regenerative, i.e. there exists $A \in \mathcal{E}\otimes\mathcal{E}$ such that $\psi(A) > 0$ and

$$\forall (x,y) \in A,\ \forall (x',y') \in A,\ \forall F \in \mathcal{E}\otimes\mathcal{E}, \quad P\big((x,y);F\big) = P\big((x',y');F\big).$$

Such a set A is called an accessible atom for the Markov chain $(Y_n)_{n\geq 1}$. Notice that the assumption of Harris-recurrence is equivalent to assuming that $\mathbb{P}_{(x,y)}((Y_n)_{n\geq 1} \in A\ i.o.) = 1$ for all $(x,y) \in E\times E$ (cf. [MT93]). The hitting time to A and the successive return times to A are defined as follows

$$T_A(1) := \inf\{n \geq 1; Y_n \in A\} \quad \text{and} \quad \forall j \geq 2,\ T_A(j) := \inf\{n > T_A(j-1); Y_n \in A\},$$

and we also define the number of visits of $(Y_n)_{n\geq 1}$ before time $n \geq 1$ to the set A by $l_n := \sum_{i=1}^n \mathbf{1}_A(Y_i)$.

Then, since $S_n = \sum_{i=1}^n \xi(Y_i)$ for all $n \geq 1$, we have:

$$\begin{aligned} S_n &= \sum_{i=1}^{T_A(1)} \xi(Y_i) + \sum_{j=1}^{l_n-1}\left[\sum_{i=1+T_A(j)}^{T_A(j+1)} \xi(Y_i)\right] + \sum_{i=1+T_A(l_n)}^{n} \xi(Y_i) \\ &= \sum_{i=1}^{T_A(1)} \xi(Y_i) + \sum_{j=1}^{l_n-1} \widetilde{\xi}(\mathcal{B}_j) + \sum_{i=1+T_A(l_n)}^{n} \xi(Y_i), \end{aligned}$$

where the blocks of observations between consecutive visits to the atom A are denoted by $\mathcal{B}_j := (Y_{1+T_A(j)}, \ldots, Y_{T_A(j+1)})$, and where $\widetilde{\xi}(\mathcal{B}_j) := \sum_{i=1+T_A(j)}^{T_A(j+1)} \xi(Y_i)$.

Introduce $\tau_A(j) := T_A(j) - T_A(j-1)$ for all $j \geq 2$. Let us remark that $(\mathcal{B}_j, \tau_A(j+1))_{j\geq 1}$ is an i.i.d. sequence (this fact follows from the strong Markov property), and thus $(\widetilde{\xi}(\mathcal{B}_j), \tau_A(j+1))_{j\geq 1}$ is also an i.i.d. sequence.

We obtain for all Borel set B

$$\begin{aligned} \mathbb{P}_\mu\left\{\frac{S_n}{\sqrt{n}} \in B\right\} &= \sum_{a,b,c=0}^{n} \mathbb{P}_\mu\Bigg\{\frac{\sum_{i=1}^a \xi(Y_i) + \sum_{j=1}^{b-1}\widetilde{\xi}(\mathcal{B}_j) + \sum_{i=n-c+1}^n \xi(Y_i)}{\sqrt{n}} \in B, \\ &\qquad\qquad T_A(1) = a, \sum_{j=2}^{b} \tau_A(j) = n-a-c, \tau_A(b+1) > c\Bigg\} \\ &= \sum_{a,b,c=0}^{n} \mathbb{P}_\mu\left\{Z_{1,a} + Z_{2,b} + Z_{3,c} \in \frac{\sqrt{n}}{\sqrt{b-1}} B\right\} \end{aligned}$$

where $Z_{1,a}$, $Z_{2,b}$ and $Z_{3,c}$ are independent and distributed according to the following measures

$$\begin{aligned}
\mathbb{P}_\mu\Big\{Z_{1,a}\in B\Big\} &:= \mathbb{P}_\mu\Big\{\frac{1}{\sqrt{b-1}}\sum_{i=1}^{a}\xi(Y_i)\in B, T_A(1)=a\Big\}\\
\mathbb{P}_\mu\Big\{Z_{2,b}\in B\Big\} &:= \mathbb{P}_A\Big\{\frac{1}{\sqrt{b-1}}\sum_{j=1}^{b-1}\widetilde{\xi}(\mathcal{B}_j)\in B, \sum_{j=2}^{b}\tau_A(j)=n-a-c\Big\}\\
\mathbb{P}_\mu\Big\{Z_{3,c}\in B\Big\} &:= \mathbb{P}_A\Big\{\frac{1}{\sqrt{b-1}}\sum_{i=1}^{c}\xi(Y_i)\in B, \tau_A(2)>c\Big\}.
\end{aligned}$$

Assume that we can state the asymptotic negligibility of $Z_{1,a}$ and $Z_{3,c}$. Then it would only remain to control the behavior of $Z_{2,b}$. Let us notice that $(\widetilde{\xi}(\mathcal{B}_j), \tau_A(j+1))_{j\geq 1}$ is a two-dimensional i.i.d. sequence which is lattice in one component. Although this sequence is lattice in one component, we can establish a limit theorem associated to $Z_{2,b}$ when $b\to+\infty$ by applying the results of [Dub82, Dub84a, Dub84b] as in [BC04, lem. 6.5]. Then, at a first glance, considering the sequence $(Y_n)_{n\geq 1}$ instead of the sequence $(X_n)_{n\geq 0}$ seems helpful.

However, assuming the existence of an atom for the double sequence $(Y_n)_{n\geq 1}$ induces very restrictive conditions on the model:

Lemma 10. *If $(Y_n)_{n\geq 1}$ possesses an accessible atom A, then there exists a state $b\in E$ such that $A := A_1\times\{b\}$ for some $A_1\in\mathcal{E}$. Furthermore, whatever the initial state $y\in E$, the chain $(X_n)_{n\geq 0}$ visits $\{b\}$ with strictly positive probability, that is:*

$$\forall y\in E,\quad \sum_{n=1}^{+\infty}Q^n\big(y,\{b\}\big)>0.$$

Proof of Lemma 10. We obviously have

$$\forall (x,y)\in E\times E,\ \forall\, b\in E,\quad P\big((x,y);\{b\}\times E\big)=\delta_{y,b}.$$

Consequently, if (x,y) and (x',y') are two elements of the atom A, then we have $y=y'$. In other words we have $A=A_1\times\{b\}$ for some $A_1\in\mathcal{E}$ and $b\in E$. Besides we obtain for all $(x,y)\in E\times E$

$$\sum_{n=1}^{+\infty}Q^n\big(y,\{b\}\big)=\sum_{n=2}^{+\infty}\mathbb{P}_{(x,y)}\big\{(X_{n-1},X_n)\in E\times\{b\}\big\}\geq\sum_{n=2}^{+\infty}\mathbb{P}_{(x,y)}\big\{Y_n\in A\big\}>0,$$

since $\psi(A)>0$ and $(Y_n)_{n\geq 1}$ is ψ-irreducible. □

B.1.2 Discussion under a minorization condition for $(Y_n)_{n\geq 1}$

To avoid the assumption of the existence of an accessible atom (and hence the strong conditions of Lemma 10), we may use the regeneration method constructed via the splitting technique under a minorization condition. However in the same way, supposing that P satisfies a minorization condition induces also very restrictive conditions on the model.

Under the ψ-irreducibility assumption, we suppose that $(Y_n)_{n\geq 1}$ is Harris-recurrent and satisfies a minorization condition, i.e. there exist a measurable function $0 \leq h < 1$ on $E \times E$ and a positive measure ν on $(E \times E, \mathcal{E} \otimes \mathcal{E})$ such that $\nu(h) > 0$ and:

$$\forall (x, y) \in E \times E,\ \forall F \in \mathcal{E} \otimes \mathcal{E}, \quad P\big((x, y); F\big) \geq h(x, y)\, \nu(F). \tag{42}$$

Lemma 11. *The Markov kernel P of $(Y_n)_{n\geq 1}$ satisfies a minorization condition if and only if there exists $b \in E$ such that $Q(b, \{b\}) > 0$.*

Proof of Lemma 11. Assume that there exists $b \in E$ such that $Q(b, \{b\}) > 0$. Define $h(x, y) := \mathbf{1}_{\{(b,b)\}}(x, y)$ and $\nu := Q(b, \{b\})\, \delta_{(b,b)}$. Let $F \in \mathcal{E} \otimes \mathcal{E}$. If $(b, b) \notin F$, then $\nu(F) = 0$ and $P\big((b, b); F\big) \geq 0$. If $(b, b) \in F$, then

$$P\big((b, b); F\big) \geq P\big((b, b); \{(b, b)\}\big) = Q(b, \{b\}) = \nu(F).$$

Hence the desired minorization condition holds.
Conversely, assume that P satisfies a minorization condition. At a first stage, assume that

$$\exists\, b \in E \ \text{ such that } \ [\ h(b, y) > 0 \ \Rightarrow \ y = b\ ]. \tag{43}$$

Then from $h(b, \cdot)\mathbf{1}_{E\setminus\{b\}}(\cdot) \equiv 0$ and (42), we obtain

$$h(b, b)Q(b, \{b\}) = \int_E h(b, z)Q(b, dz) = Ph(b, b) \geq h(b, b)\nu(h),$$

thus $Q(b, \{b\}) \geq \nu(h) > 0$, which ends the proof of $Q(b, \{b\}) > 0$. Then it only remains to prove (43): let $(a, b) \in E^2$ be such that $h(a, b) > 0$, define the following set $T := \{y \in E;\, h(b, y) > 0\}$, and let us prove that $T = \{b\}$. Let $g(x, y) := h(x, y)\mathbf{1}_{E\setminus\{b\}}(x)$, then we obtain from $g(b, \cdot) \equiv 0$ and (42):

$$0 = \int_E g(b, z)Q(b, dz) = Pg(a, b) \geq h(a, b)\nu(g),$$

thus $\nu(g) = 0$. Let us define $f(x, y) := h(x, y)\mathbf{1}_{\{b\}}(x)$. Since $\nu(h) > 0$, we obtain $\nu(f) = \nu(h) - \nu(g) > 0$, which implies in particular that T is not empty. Let $y_0 \in T$, then we obtain from $f(x, \cdot)\mathbf{1}_{E\setminus\{b\}}(x) \equiv 0$ and (42):

$$\delta_b(y_0)Pf(b, b) = \int_E f(y_0, z)Q(y_0, dz) = Pf(b, y_0) \geq h(b, y_0)\nu(f) > 0,$$

which implies $y_0 = b$, and thus (43). □

In conclusion, since $P((x, y); .)$ does not depend on $x \in E$, $(Y_n)_{n\geq 1}$ can possess an accessible atom or satisfy a minorization condition only under very restrictive conditions on $(X_n)_{n\geq 0}$ (cf. Lemmas 10 and 11). For instance, whenever $(X_n)_{n\geq 0}$ is dominated by the Lebesgue measure on $\mathbb{R}^d$ (as the AR(1) processes defined in (28)), $(Y_n)_{n\geq 1}$ cannot possess an accessible atom and cannot satisfy a minorization condition: in these quite usual models, the results of [BC11] cannot be applied to $(Y_n)_{n\geq 1}$.

B.2 Application of the results of [Jen89]

As explained just before, the application of [BC11] to the double sequence $(Y_n)_{n\geq 1}$ fails to provide Berry-Esseen theorem for many general models. That is why in this subsection, we directly deal with the simple sequence $(X_n)_{n\geq 0}$.

B.2.1 Discussion under existence of an atom for $(X_n)_{n\geq 0}$

By now, assume that $(X_n)_{n\geq 0}$ is ψ-irreducible, Harris-recurrent and that $(X_n)_{n\geq 0}$ has an accessible atom denoted by A, i.e. $\psi(A) > 0$ and

$$\forall x \in A,\ \forall x' \in A,\ \forall F \in \mathcal{E}, \quad Q\big(x; F\big) = Q\big(x'; F\big).$$

Define

$$T_A(1) := \inf\{n \geq 0; X_n \in A\} \quad \text{and} \quad \forall j \geq 2,\ T_A(j) := \inf\{n > T_A(j-1); X_n \in A\},$$

and for all $n \geq 0$, $l_n := \sum_{i=0}^{n} \mathbf{1}_A(X_i)$. Then, for all $n \geq 1$, we have:

$$S_n \;=\; \sum_{i=1}^{T_A(1)} \xi(X_{i-1}, X_i) + \sum_{j=1}^{l_n - 1} \widetilde{\xi}(\mathcal{B}_{j-1}, \mathcal{B}_j) + \sum_{i=1+T_A(l_n)}^{n} \xi(X_{i-1}, X_i),$$

where the blocks of observations between consecutive visits to the atom A are denoted by $\mathcal{B}_j := (X_{1+T_A(j)}, \ldots, X_{T_A(j+1)})$, and where $\widetilde{\xi}(\mathcal{B}_{j-1}, \mathcal{B}_j) := \sum_{i=1+T_A(j)}^{T_A(j+1)} \xi(X_{i-1}, X_i)$.

Let $\tau_A(j)$ denote $\tau_A(j) := T_A(j) - T_A(j-1)$ for all $j \geq 2$. Let us remark that although $(\mathcal{B}_j, \tau_A(j+1))_{j\geq 1}$ is an i.i.d. sequence, the sequence $(\widetilde{\xi}(\mathcal{B}_{j-1}, \mathcal{B}_j), \tau_A(j+1))_{j\geq 1}$ is no longer i.i.d. but one-dependent.

We obtain for all Borel set B

$$\mathbb{P}_\mu\left\{\frac{S_n(p)}{\sqrt{n}} \in B\right\} \;:=\; \sum_{a,b,c=0}^{n} \mathbb{P}_\mu\left\{Z_{1,a} + Z_{2,b} + Z_{3,c} \in \frac{\sqrt{n}}{\sqrt{b-1}} B\right\}$$

where $Z_{1,a}$, $Z_{2,b}$ and $Z_{3,c}$ are distributed according to the following measures

$$\begin{aligned}
\mathbb{P}_\mu\left\{Z_{1,a} \in B\right\} \;&:=\; \mathbb{P}_\mu\left\{\frac{1}{\sqrt{b-1}} \sum_{i=1}^{a} \xi(X_{i-1}, X_i) \in B, T_A(1) = a\right\} \\
\mathbb{P}_\mu\left\{Z_{2,b} \in B\right\} \;&:=\; \mathbb{P}_A\left\{\frac{1}{\sqrt{b-1}} \sum_{j=1}^{b-1} \widetilde{\xi}(\mathcal{B}_{j-1}, \mathcal{B}_j) \in B, \sum_{j=2}^{b} \tau_A(j) = n - a - c\right\} \\
\mathbb{P}_\mu\left\{Z_{3,c} \in B\right\} \;&:=\; \mathbb{P}_A\left\{\frac{1}{\sqrt{b-1}} \sum_{i=1}^{c} \xi(X_{i-1}, X_i) \in B, \tau_A(2) > c\right\}.
\end{aligned}$$

Notice that $Z_{1,a}$, $Z_{2,b}$ and $Z_{3,c}$ are no longer independent and that $(\widetilde{\xi}(\mathcal{B}_{j-1}, \mathcal{B}_j), \tau_A(j+1))_{j\geq 1}$ is a two-dimensional one-dependent sequence which is lattice in one component. To establish a limit theorem when $b \to +\infty$ in spite of this one-dependence, we can apply the usual Nagaev-Guivarc'h method, as in [Jen89, th. 2] where Jensen manages to obtain non-parametric results on bivariate functions thanks to a judicious combination between regeneration and Fourier methods.

However, this method induces an unusual Cramér assumption (this condition concerns equally the lattice v.a. $\tau_A(j)$) as well as an intricate covariance matrix (in particular, this matrix depends on the variance of $\tau_A(j)$).

Furthermore, as explained in Introduction, Jensen must assume some block moment conditions which are far from being easy to be verified, except when the initial probability μ is either concentrated at one point in the atom A (i.e. there exists $x_0 \in A$ such that $\mu = \delta_{x_0}$), or is dominated by some multiple of the stationary probability π.

B.2.2 Discussion under a minorization condition for $(X_n)_{n\geq 0}$

From now on, assume that $(X_n)_{n\geq 0}$ is ψ-irreducible, Harris-recurrent, has no accessible atom but satisfies a minorization condition, i.e there exist a measurable function $0 \leq h < 1$ on E and a positive measure ν on $(E, \mathcal{E})$ such that $\nu(h) > 0$ and:

$$\forall x \in E,\ \forall F \in \mathcal{E}, \quad Q\big(x; F\big) \geq h(x)\, \nu(F). \tag{44}$$

When assuming that the chain $(X_n)_{n\geq 0}$ only satisfies the minorization condition (44), we must construct a new Markov chain $(\check{X}_n)_{n\geq 0}$ whose state space is $E\times\{0,1\}$ and which possesses the atom $A := E \times \{1\}$ using the splitting method of Nummelin. Then we make all the preceding job with this new chain $(\check{X}_n)_{n\geq 0}$. In particular, $(\check{X}_n)_{n\geq 0}$ must satisfy some complex block moment conditions which can be easily verified only under the following strong restrictions on the initial probability $\check{\mu}$ of this split chain: the initial probability $\check{\mu}$ is either concentrated at one point in $E \times \{1\}$, or dominated by some multiple of the stationary probability of the split chain $(\check{X}_n)_{n\geq 0}$.

Let us recall that $\check{\mu}$ is defined as follows:

$$\forall F \in \mathcal{E}, \qquad \check{\mu}((F,i)) := \begin{cases} \int_F \ (1-h(x))\ \mu(dx) & \text{if} \quad i = 0 \\ \\ \int_F \ h(x)\ \mu(dx) & \text{if} \quad i = 1. \end{cases}$$

Then, it is easy to see [5] that, whatever $x_0 \in E$, there exist no probability measure μ for the chain $(X_n)_{n\geq 0}$ which can match with $\check{\mu} := \delta_{\{x_0\}\times\{1\}}$ for the split chain $(\check{X}_n)_{n\geq 0}$. In other words, applying regeneration method with explicit moment condition whenever the chain only satisfies a minorization condition imposes the domination of the initial probability μ by some multiple of the stationary probability π.

In conclusion, even in non-parametric cases, the regeneration method is not as efficient as our method. Unlike regeneration results, our results are obtained under weak conditions on the

[5]Indeed, if $\check{\mu}(E \times \{0\}) = 0$, then $h = 1$ μ-a.s, which is in contradiction with the fact that $h < 1$.

initial probability μ of the chain $(X_n)_{n\geq 0}$. In particular, they hold true in the simple case where μ is a Dirac distribution at $x_0 \in E$. Furthermore, the effective control of constants in the regenerative method probably still needs much work to be done (even for univariate theorem, cf. [BC11]).

References

[Bal00] V. Baladi. *Positive Transfer Operators and Decay of Correlations*, volume 16 of *Adv. Ser. Nonlinear Dynam.* World Scientific, 2000.

[BC04] P. Bertail and S. Clémençon. Edgeworth expansions of suitably normalized sample mean statistics for atomic Markov chains. *Probab. Theory Relat. Fields*, 130:388–414, 2004.

[BC11] P. Bertail and S. Clémençon. A renewal approach to markovian U-statistics. *Math. Methods Statist.*, 20:79–105, 2011.

[Bil61] P. Billingsley. *Statistical inference for Markov Processes.* The University of Chicago Press, 1961.

[Bol82] E. Bolthausen. The Berry Esseen theorem for strongly mixing Harris recurrent Markov chains. *Probab. Theory Related Fields*, 60:283–289, 1982.

[Dah89] R. Dahlhaus. Efficient parameter estimation for self similar processes. *Ann. Statist.*, 17:1749–1766, 1989.

[DM06] A. Douc, R. ; Guillin and E. Moulines. Bounds on regeneration times and limit theorems for subgeometric Markov chains. *Ann. Inst. H. Poincaré Probab. et Statist*, 44:239–257, 2006.

[Dub82] I. Dubinskaite. Limit theorems in $\mathbb{R}^k$, i. *Lithuanian Math. J.*, 22:129–140, 1982.

[Dub84a] I. Dubinskaite. Limit theorems in $\mathbb{R}^k$, ii. *Lithuanian Math. J.*, 24:256–265, 1984.

[Dub84b] I. Dubinskaite. Limit theorems in $\mathbb{R}^k$, iii. *Lithuanian Math. J.*, 24:325–334, 1984.

[Dur80] J. Durbin. Approximations for densities of sufficient estimators. *Biometrika*, 67:311–333, 1980.

[DY07] D. Dehay and J.F. Yao. On likelihood estimation for discretely observed Markov jump processes. *Aust. N. Z. J. Stat.*, 49:93–107, 2007.

[Fel71] W. Feller. *An introduction to probability theory and its applications, Vol. II.* John Wiley and Sons, New York, 1971.

[Fer] D. Ferré. Synthesis document on Keller Liverani theorem. available at http://www.insa-rennes.fr/deborah-ferre.

[Fer10] D. Ferré. Développement d'Edgeworth d'ordre 1 pour des M-estimateurs dans le cas de chaînes V-géométriquement ergodiques. *CRAS*, 348:331–334, 2010.

[FHL] D. Ferré, L. Hervé, and J. Ledoux. Limit theorems for stationary Markov processes with L^2-spectral gap. Accepted in Ann. Henri Poincaré.

[Fuh06] C.-D. Fuh. Efficient likelihood estimation in state space models. *Ann. Statist.*, 34:2026–2068, 2006.

[Gän72] P. Gänssler. Note on minimum contrast estimators for Markov processes. *Metrika*, 19:115–130, 1972.

[GH94] F. Götze and C. Hipp. Asymptotic distribution of statistics in time series. *Ann. Statist.*, 22:2062–2088, 1994.

[HH01] H. Hennion and L. Hervé. *Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness.* Lect. Notes Math. Springer, 2001.

[HLP] L. Hervé, J. Ledoux, and V. Patilea. A Berry-Esseen theorem on M-estimators for V-geometrically ergodic Markov chains. Accepted in Bernoulli.

[HP10] L. Hervé and F. Pène. The Nagaev-Guivarc'h method via the Keller-Liverani theorem. *Bull. Soc. Math. France*, 138 (3):415–489, 2010. Ann. Inst. H. Poincaré Probab. Statist.

[Jen89] J. L. Jensen. Asymptotic expansions for strongly mixing Harris recurrent Markov chains. *Scand. J. Statist.*, 16:47–63, 1989.

[KL99] G. Keller and C. Liverani. Stability of the spectrum for transfer operators. *Ann. Sc. Norm. Super. Pisa Cl. Sci (4)*, XXVIII:141–152, 1999.

[Liv04] C. Liverani. Invariant measure and their properties. a functional analytic point of view. In Ann. Sc. Norm. Super. Pisa, editor, *Part II: Topological Geometrical and Ergodic Properties of Dynamics*, CCRM, 2004.

[LRZ03] O. Lieberman, J. Rousseau, and D.M. Zucker. Valid asymptotic expansions for the maximum likelihood estimator of the parameter of a stationary, Gaussian, strongly dependent process. *Ann. Statist.*, 31:586–612, 2003.

[Mal87] V.K. Malinovskii. Limit theorems for Harris Markov chains I. *Theory Prob. Appl.*, 31:269–285, 1987.

[MT93] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability.* Springer Verlag, New-York, Heidelberg, Berlin, 1993.

[Pfa73] J. Pfanzagl. Asymptotic expansions related to minimum contrast estimators. *Ann. Statist.*, 1(6):993–1026, 1973.

[Rao72] B. L. S. Prakasa Rao. Maximum likelihood estimation for Markov processes. *Ann. Inst. Statist. Math.*, 24:333–345, 1972.

[Rou65] G.G. Roussas. Asymptotic inferencer in Markov process. *Ann. Math. Statist.*, 36:978–982, 1965.

[RW97] F. Räbiger and M.P. H. Wolff. On the approximation of positive operators and the behaviour of the spectra of the approximants. *Integral Equations Operator Theory*, 28(1):72–86, 1997.

[Smi55] W. L. Smith. Regenerative stochastic processes. *Proc. Roy. Soc. London. Ser. A.*, 232:6–31, 1955.